The Philosophical Magazine

FIRST PUBLISHED IN 1798

A Journal of Theoretical Experimental and Applied Physics

Vol. 4 January 1959 No. 37
Eighth Series

£1 5s. 0d., plus postage
Annual Subscription £13 10s. 0d., payable in advance

Printed and Published by
TAYLOR & FRANCIS LTD
RED LION COURT, FLEET STREET, LONDON, E.C.4

THE PHILOSOPHICAL MAGAZINE

Authors wishing to submit papers for publication in the Journal should send manuscripts directly to the Publishers.

Manuscripts should be typed in *double* spacing on one side of quarto (8×10 in.) paper, and authors are urged to aim at absolute clarity of meaning and an attractive presentation of their texts.

References should be listed at the end in alphabetical order of authors and should be cited in the text in terms of author's name and date. Diagrams should normally be in Indian ink on white card, with lettering in soft pencil, the captions being typed on a separate sheet.

A leaflet giving detailed instructions to authors on the preparation of papers is available on request from the Publishers.

Authors are entitled to receive 25 offprints of a paper in the Journal free of charge, and additional offprints can be obtained from the Publishers.

The *Philosophical Magazine* and its companion journal, *Advances in Physics*, will accept papers for publication in experimental and theoretical physics. The *Philosophical Magazine* publishes contributions describing new results, letters to the editor and book reviews. *Advances in Physics* publishes articles surveying the present state of knowledge in any branch of the science in which recent progress has been made. The editors welcome contributions from overseas as well as from the United Kingdom, and papers may be published in English, French and German.

A HISTORY OF MATHEMATICS

from antiquity to the early nineteenth century

by J. F. SCOTT, B.A., D.Sc., Ph.D.
Vice-Principal of St. Mary's College, Strawberry Hill, Twickenham, Middlesex

Author of *The Scientific Work of René Descartes* (1596–1650), *Mathematical Work of John Wallis, D.D., F.R.S.* (1616–1703), and other works.

CONTENTS : Mathematics in Antiquity—Greek Mathematics—The Invention of Trigonometry—Decline of Alexandrian Science and the Revival in Europe—Mathematics in the Orient—Progress of Mathematics during the Renaissance—New Methods in Geometry—The Rise of Mechanics—the Invention of Decimal Fractions and of Logarithms—Newton and the Calculus—Taylor and Maclaurin, the Bernoullis and Euler, Related Advances—The Calculus of Variations, Probability, Projective Geometry, Non-Euclidean Geometry—Theory of Numbers—Lagrange, Legendre, Laplace, Gauss.

This volume is intended primarily to help students who desire to have a knowledge of the development of the subject but who have too little leisure to consult original works and documents. The author has availed himself of the facilities afforded by the Royal Society and other learned Societies to reproduce extracts from manuscripts and many scarce works.

$9\frac{3}{4} \times 6\frac{3}{4}$ Price 3 guineas

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

THE SCIENTIFIC WORK OF RENÉ DESCARTES 1596-1650

BY
J. F. SCOTT, B.A., M.SC., PH.D.

With a Foreword by
H. W. TURNBULL, M.A., F.R.S.

This book puts the chief mathematical and physical discoveries of Descartes in an accessible form and fills an outstanding gap upon the shelf devoted to the history of philosophy and science. There is to be found in this volume the considerable contribution that Descartes made to the physical sciences, which involved much accurate work in geometrical optics and its bearing upon the practical problem of fashioning lenses, as also the deeper problems of light and sight and colour. The careful treatment that Dr. Scott has accorded to the work of Descartes is very welcome. The book is well worth reading and will be an asset to all libraries. This publication is recommended and approved by the Publication Fund Committee of the University of London.

212 *pp.* 7×10 *amply illustrated*

PRICE £1–0–0 NET

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Journal of Electronics and Control

A Philosophical Magazine Associated Journal

Contents of November, 1958

Price per part £1 5s. plus postage
Price per volume £7 post free, payable in advance
6 monthly issues per volume

Printed and Published by

TAYLOR & FRANCIS LTD

RED LION COURT, FLEET STREET, LONDON, E.C.4

Orders originating in U.S.A. and Canada should be sent to the Academic Press Inc., 111 Fifth Avenue, New York, 3, N.Y., U.S.A.

Contents of November, 1958

Advances in Physics

A Quarterly Supplement of the Philosophical Magazine

Contents of October, 1958

Price per part £1 plus postage

Price per annum £3 15s. post free

Printed and Published by

TAYLOR & FRANCIS LTD

RED LION COURT, FLEET STREET, LONDON, E.C.4

Physics in Medicine and Biology

A Taylor & Francis Journal published in association with The Hospital Physicists' Association

Contents of October, 1958

Subscription price per volume £3 10s. post free, payable in advance

4 parts per volume—£1 per part plus postage

Printed and Published by

TAYLOR & FRANCIS, LTD

RED LION COURT, FLEET STREET, LONDON, E.C.4

Orders originating in U.S.A. and Canada should be sent to the Academic Press Inc., 111 Fifth Avenue, New York, 3, N.Y., U.S.A.

Contents of October, 1958

Price per part £1 5s. 0d. plus postage
Subscription per volume (4 issues) £4 15s. 0d. post free, payable in advance

Printed and Published by
TAYLOR & FRANCIS LTD
RED LION COURT, FLEET STREET, LONDON, E.C.4

Orders originating in U.S.A. and Canada should be sent to the Academic Press Inc., 111 Fifth Avenue, New York, 3, N.Y., U.S.A.

THE

PHILOSOPHICAL MAGAZINE

A JOURNAL OF THEORETICAL, EXPERIMENTAL AND APPLIED PHYSICS

First Published in 1798

[EIGHTH SERIES—VOL. 4]

Experimental Investigations of Exciton Spectra in Ionic Crystals†

By S. Nikitine

Laboratoire de Spectroscopie et d'Optique du Corps Solide, Institut de Physique, Université de Strasbourg

[Received February 4, 1958 ; and in revised form July 8, 1958]

Abstract

An account is given of the experimental results which have been obtained since 1953 in the author's laboratory for the spectroscopy of crystals. It has been shown that, with a number of crystals, including Cu_2O, CuCl, CuBr, CuI, AgI, HgI_2, PbI_2, TlCl, TlBr, and TlI, line spectra can be observed at low temperatures in absorption and in emission. The reflection spectra have also been studied. The occurrence of a line spectrum appears to be a characteristic feature of pure ionic crystals and although the subject is still under investigation it is now possible to outline the general properties of these spectra.

Three classes of line spectra have been observed. The spectra of two of the classes are in reasonable agreement with the predictions of recent theories of exciton spectra while those of the third class are most probably also exciton spectra of a type which has not yet been treated theoretically. Other interpretations are, however, possible. When very thin samples ($0{\cdot}1\ \mu$) are used for the observation of spectra of the first class, there is only a weak continuum and the spectra consist mainly of a few strong characteristic lines (' raies ultimes ') ; these might possibly be used for analytical purposes. The intensities of lines in the spectra of the third class are extremely sensitive to the quality of the crystal. The results of our experiments are in qualitative agreement with recent theories of exciton spectra. It will, however, be necessary to extend both the experimental and the theoretical work before an entirely satisfactory interpretation of the spectra of solids can be given.

§ 1. Introduction

Much fundamental information about the energy levels of atoms and molecules has been obtained from spectroscopic investigations and there is no reason why the spectroscopy of solids should not furnish equally important information on the energy levels of crystals. Very thin specimens must be used because the absorption coefficient of a pure crystal may

† Communicated by the Author.

be of the order of 10^6 cm^{-1}. These thin sheets must, moreover, have the compact structure of large crystals and the spectra must be investigated at very low temperatures. Because of these experimental difficulties very few reliable observations on the spectra of pure crystals were made until quite recently.

The purpose of this paper is to review the experimental work on the spectroscopy of pure solids, including both insulators and intrinsic semiconductors, which has been carried out in the author's laboratory since 1953.

§ 2. Theoretical Background

2.1. *The Internal Photoeffect and the Exciton*

Only a very general outline of the theory will be presented, since it is not yet adequate for the interpretation of all the available experimental results. Consider a pure monatomic crystal which may be an insulator or an intrinsic semiconductor. When an electron is transferred from the valence band to the conduction band after the absorption of a photon of suitable energy, both the electron and the hole left in the valence band may be mobile and give rise to photoconductivity. If excitons did not exist, a crystal should have a continuous absorption band with a more or less sharp long-wavelength limit, corresponding to the limit of the photoconductive response curve. The photoconductive response curve is, however, frequently displaced to shorter wavelengths compared with the absorption curve and it has therefore been recognized that crystals may absorb by a mechanism other than that described above.

From an atomic point of view, the photoconductive effect arises from transitions of electrons from bound to free states. Frenkel (1931 a, b) proposed that electrons could also exist in solids in bound excited states similar to atomic excited states. In a crystal, the excitation can travel from one atom to another and he called this mobile state of excitation an 'exciton'.

2.2. *Exciton Spectra*

Wannier (1937, see also, Mott and Gurney 1948) has shown that the energy of an exciton can be divided into two parts, its kinetic energy, and the binding energy of the electron to the hole. This energy is quantized as in a hydrogen-like atom, if the potential energy tends to zero as $1/r$ for large distances. It can be shown that, in the allowed optical transitions, the crystal goes from an unexcited state to one of these quantized exciton states. This should give a spectrum of sharp lines forming hydrogen-like series. According to Peierls (1932), electronic transitions may combine with vibrational transitions and, in particular, with optical vibrations. It might therefore be anticipated that the exciton spectra would consist of sharp lines, forming hydrogen-like series repeated sometimes as vibrational satellites. The energy levels of the excitons are in fact broadened into excitation bands on account of the translational energy. The selection rules are discussed below.

The energies of the quantized exciton states must be calculated before the spectra can be predicted and this has been done by Mott using an elegant though rough approximation. Other calculations have been made in which more accurate and elaborate approaches to the problem have been tried (Heller and Marcus 1951, Dexter 1951, Dykman and Pekar 1952, Overhauser 1956, Muto and Okuno 1956, Haken 1956 a, b, c, d, e, 1957, Meyer 1956, Dresselhaus 1956, 1957 and Elliott 1957). The results of some of these calculations will be summarized in § 2.3.

The excited electron is under the action of the Coulomb field of the hole and of the local field in the crystal. Mott suggests the introduction for the excited electron of a potential energy of the form $-e^2/\kappa r$, where κ, the optical dielectric constant, takes care of the mean influence of the surrounding crystal, apart from the hole, on the excited electron. This can only be a rough approximation for orbits of small radius; but it is likely to be a good approximation at large distances from the hole. According to the theory, the radii of the orbits of the exciton are related to the corresponding orbits of the hydrogen atom by the equation $r_{ex} = \kappa r_H$. As κ can be of the order of 10, it appears that Mott's approximation will be good for all but the first orbit.

With Mott's expression for the potential energy, it can be shown that the quantized binding energy (cm^{-1}) of the exciton in the crystal is

$$E_k/ch = E_0/ch - R'/k^2, \qquad R' = R\mu/\kappa^2 m \qquad . \; . \; . \quad (1)$$

where E_0 is the lowest energy state of the conduction band, corresponding to an unbound electron and hole. R is Rydberg's constant; h, Planck's constant; c, the velocity of light; μ the reduced mass of the hole-electron system; m, the mass of the electron at rest and k, a quantum number. This model of the exciton is often compared with that of positronium. Some authors differentiate between the exciton model of Mott and that of Frenkel, who seemed to envisage a hydrogen-like model, in which the mass of the hole was large compared with that of the electron. In Mott's model, both the electron and the hole have comparable masses and should describe orbits of comparable radii (in terms of Bohr's model for the hydrogen atom with precession).

The predicted exciton spectrum corresponds to the jump of an electron from an occupied level of the crystal to an exciton level, the energy of which is given by (1). If the energy of the normal state of the crystal is E_a, the energy of formation of an exciton is $E_k - E_a$. The value of E_a is, however, not known and the spectrum of the exciton has therefore to be represented in the following way:

$$\nu_k = \nu_\infty - R'/k^2; \; \nu_\infty = (E_0 - E_a)/hc \qquad . \; . \; . \; . \; . \quad (2)$$

ν_∞ is the limit of the series of hydrogen-like lines expected in the spectrum of the exciton. For $\nu > \nu_\infty$, the spectrum should be continuous; ν_∞ can, however, not yet be derived from the theory. R' is difficult to evaluate and there is uncertainty as to which value of κ and μ should be used. These matters have been examined recently and will be discussed in § 2.3. Photoconductivity might be expected to occur in the continuous

part of the spectrum where $\nu > \nu_\infty$. For $\nu < \nu_\infty$, no photoconductive effect should be observed at very low temperatures as the electron is transferred to a bound excited state.

According to Peierls (1932), a vibrational term should be added to the above formula. This does not appear in a simple form in recent theoretical papers and would presumably lead to complicated calculations (Haken 1957). With the vibrational term, the complete formula is:

$$\nu_k = \nu_\infty - R'/k^2 \pm p\nu_{1R} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where ν_{1R} is the infra-red frequency of the crystal and p is an integer. As $R' \ll R$ on account of the high value of κ^2 the electronic term should often be smaller than the vibrational term and one would expect the spectrum of the exciton to consist of a repetition of hydrogen-like series of absorption lines, the limits of which should be separated by ν_{1R}. R' need not necessarily be the same in the different series.

Mott's theory has recently extended to uniaxial crystals (Barriol *et al.* 1956). Assuming the effective mass to be constant and the anisotropy of κ to be small, it can be shown that the degeneracy of the exciton levels is destroyed and the exciton lines should split into multiplets polarized at right angles. The exciton spectra can then become quite complicated; in anisotropic crystals, the effective mass can be anisotropic and dichroic absorption can occur.

The line structure of the spectra of solids should be observed only at very low temperatures because the closely spaced lines will be broadened by thermal vibrations at higher temperatures. The effect of temperature, according to a recent theory of Elliott (1957), discussed below, is to stimulate a background absorption between the lines rather than to influence the breadth of the lines. These factors account for that part of the spectrum which is continuous at high temperatures and does not give rise to a photoconductive effect.

2.3. *Recent Advances in the Theory*

In the previous section, a general account of the theory of exciton spectra has been given. A detailed treatment of the more rigorous theories has been omitted, because the theory of Mott leads to the same results and is sufficient for our purposes. Important advances in the theory have, however, recently been made by Haken (1956 a, b, c, d, e, 1957), Meyer (1956), Dresselhaus (1956, 1957), and particularly by Elliott (1957). These developments seem to open the way to a better understanding of exciton spectra and the essential results will be summarized in this section.

2.3.1. *The exciton and the polarization of the lattice*

Haken (1956 a, b, c, d, e) has made use of the theory of the polaron to develop a theory of the exciton. He has been able to take into account the polarization of crystal by the hole and the electron forming the exciton. An excellent review of the theory of the exciton has recently been given by Haken (1957) to whom the author is indebted for the communication of a manuscript of the paper and for numerous stimulating discussions.

According to the theory of Haken, the interaction potential in Schrödinger's equation is of the form:

$$\frac{-e^2}{\kappa_0 r_{12}} + \frac{e^2}{r_{12}}\left(\frac{1}{\kappa_0} - \frac{1}{\kappa_s}\right)[1 - \tfrac{1}{2}\{\exp(-u_1 r_{12}) + \exp(-u_2 r_{12})\}] \qquad (4)$$

where r_{12} is the distance between the electron and the hole, κ_0 the optical and κ_s the static dielectric constant, and $u_i = \sqrt{(2m_i^*\omega/h)}$. m_i^* is the effective mass of the hole or the electron and ω the frequency of the longitudinal optical mode of the crystal; the discussion of this formula shows that for $r_{12} > u_i^{-1}$ (large orbits) the potential reduces to $-e_2^2/\kappa_s r_{12}$; for $r_{12} < u_i^{-1}$ (small orbits) the potential is approximately $-e^2/\kappa_0 r_{12}$.

The important result of this theory is therefore that, in Mott's formula, κ may take values between κ_s and κ_0 for different orbits of the exciton. It is, however, likely that in most cases the value κ_0 has to be taken for the first and possibly the second orbit; the other orbits correspond to the value κ_s. The lines of series consequently do not necessarily obey a hydrogenic law. The author was led by the experimental data to suggest in his early publications that κ in Mott's expression might take different values for different orbits. A further consequence of the approximation which includes the polarization of the crystal by the hole and the electron is that the effective mass of the exciton may vary with the orbit. Dresselhaus (1956), who did not take polarization into account, found that the effective mass of the exciton had the same value in all exciton states. When the polarization of the surrounding crystal is introduced into the theory, it is found that the effective mass varies with the exciton state.

2.3.2. *Transition probabilities for exciton states*

The transition probabilities from the ground state of the crystal to exciton states have been studied recently by Dresselhaus (1957), and by Haken (1957) and Elliott (1957) to both of whom the author is greatly obliged for stimulating discussions and correspondence on the subject as well as for the communication of results prior to publication.

Two classes of exciton spectra are predicted by these theories:

(i) The spectra of the first class result from direct transitions in which the vector $\mathbf{k} = 0$. In this class, only transitions to S exciton states are allowed. The spectrum then consists of a hydrogenic series of lines, the first line corresponding to $k = 1$ in Mott's equation. The oscillator strengths of the lines in the series are of the order $(k \,.\, \kappa)^{-3}$. These spectra should therefore begin with a very strong line, followed by a second line ($k = 2$) with an intensity only 0·12 times the first.

(ii) The spectra of the second class, the existence of which was first recognized by Elliott (1957), belong to forbidden transitions corresponding to a small value of the vector $\mathbf{k}$. Elliott also calculated the form of the continuous spectrum at the limit of the series. In this class of spectra, the only transitions allowed are to P exciton states. The spectrum should consist of a hydrogenic series of lines with the first line corresponding to $k = 2$ in Mott's equation. The intensities in this class of spectra are

much smaller than in the first class. The f value of the first line is of the order of $0{\cdot}03\,\kappa^{-5}$ and the f values of the succeeding lines vary as $(k^2-1)\,.\,k^{-5}$; the intensity of the second line of the series is therefore 0·4 times that of the first line.

It should not be difficult to separate the two classes of spectra experimentally. The spectra of the first class should be observed only with very thin samples. The first line is much stronger than the succeeding lines in the series, so that it may sometimes be the only one which appears. If there are several lines, the second line will correspond to $k=2$ in Mott's equation. The first line, as shown by Haken, may not be suitably represented by Mott's equation with the same value of κ. The spectra of the second class should be found with thicker specimens. The lines in the series are of comparable intensity so that several lines should occur and the first line of the series must correspond to $k=2$ in Mott's formula.

§ 3. Experimental Methods

3.1. *Early Experimental Work*

Exciton lines have seldom been observed even though experimental work on solid state spectroscopy has been carried out for many years. A considerable number of the observations have been made at room temperature and at 77·3°K and the exciton lines are usually observed only at lower temperatures. In much experimental work, good thermal contact may not have been established between the crystal and the cryostat so that the temperature is not known accurately. Finally, it has already been mentioned that crystalline films for investigations by transmission have to be very thin (between 10^{-4} and 10^{-6} cm). Films with these thicknesses can easily be produced by evaporation, but a compact crystalline film is not obtained unless the material is condensed on a glass or a quartz substrate heated to a sufficiently high temperature or else the film is heat treated after deposition. Such films alone show an absorption spectrum with a well developed structure. This important experimental detail which was reported by Zollweg (1955) has been established independently in the author's laboratory. These precautions were not observed in early work and the films which were used may not have had a sufficiently compact structure. Spectrographic equipment with rather high dispersion is required for the detection of the narrow lines.

The early results are unlikely to have been obtained with compact crystals, and there is little doubt that many experimental features were overlooked. The important contribution of the Göttingen school of R. W. Pohl lay in the establishment of the general characteristics of the absorption spectra of a wide range of crystalline solids (Hilsch and Pohl 1928, Fesefeldt and Gyulai 1929, Fesefeldt 1930, 1931, see also. Schneider and O'Bryan 1937). The absorption spectra of the alkali halides observed by Hilsch and Pohl are generally attributed to the formation of excitons. In the light of our results with other substances, we believe that the breadth of the absorption bands which they observed is at least an order of magnitude larger than the value which would be associated with exciton

lines ; a careful investigation of these spectra with higher resolution may show a line structure in or near the bands observed by Hilsch and Pohl (1928) as observed by Martienssen (1957). It will be observed that line spectra have frequently been found in regions where either smooth absorption curves or very faint fluctuations in absorption have been reported in earlier papers. An extensive bibliography with an atlas of the absorption spectra of crystals is given in Landolt-Börnstein (1955).

3.2. *Experimental Techniques*

Our experimental technique is simple and conventional. Until recently, we have been mainly interested in the observation of narrow lines in the absorption spectra of crystals and in the determination of the wavelengths of these lines, but photometric studies of the lines and of the continuous spectra are now being undertaken. The difficulties which arise in the preparation of thin compact films can be overcome only in rare cases and films prepared from compressed powders do not give satisfactory results. Reflection spectra may provide important information and residual and ' missing ' rays have been observed in this way (Nikitine and Reiss 1956 a). These spectra, although obviously related to exciton spectra, belong to another part of the problem. The observations on these spectra, which are being studied in the author's laboratory at the present time, will be published elsewhere. Transmission spectroscopy with compact films has so far been the only successful method for the investigation of exciton spectra.

Thin films for our studies of absorption spectra are prepared in different ways:

(i) Fused substances are compressed between two Pyrex or quartz plates and solidified.

(ii) Whenever possible, thin single crystals are grown if these are required.

(iii) Evaporated films are deposited in vacuum on a heated Pyrex or quartz plate.

(iv) Evaporated metallic films are attacked with halogen vapour or with oxygen.

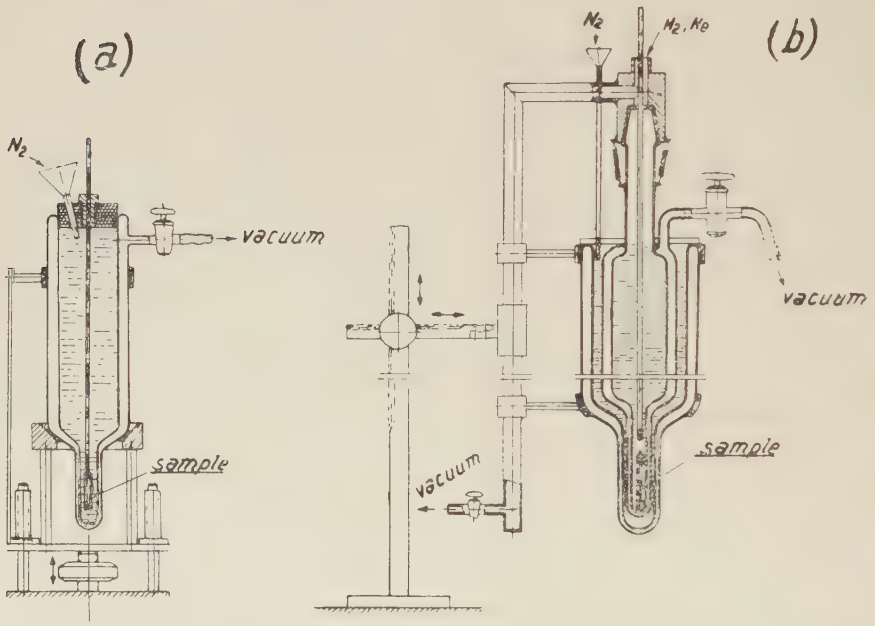
(v) In some exceptional cases, in which the absorption or the background is comparatively low, crystal sheets of about 10 or 15 μ are prepared by grinding.

When specimens of greater thickness can be employed, as with cuprous oxide and cadmium sulphide, grinding probably provides the most reliable technique. Samples prepared according to methods (i), (iii) and (iv) have been studied both before and after annealing. It is as yet too early to assess the advantages of the different methods, all of which are of considerable use.

The spectroscopic arrangement is very simple. Light from a Philips tungsten ribbon filament lamp is focused on the sample and the

transmitted light is focused on the slit of a spectrograph. The Huet AII and Zeiss three-prism spectrographs are used in the visible, and the QU 24 Zeiss quartz spectrograph in the ultra-violet. The spectra are recorded photographically. The cryostats used with liquid nitrogen, liquid hydrogen and liquid helium are of the conventional type which is in service at the laboratory of Bellevue. One of the cryostats used for liquid helium is shown in fig. 1 (*b*). The cryostat used for measurements in the ultra-violet was of fused quartz and the lower parts of the cryostats were never silvered. Simpler cryostats of similar type were used with liquid nitrogen (fig. 1 (*a*)). Great care has been taken in the establishment

Fig. 1



Cryostats used (*a*) For liquid nitrogen. (*b*) For liquid hydrogen and liquid helium.

of the temperature of the sample which was usually either supported on a glass or quartz plate or pressed between two glass or quartz plates. To ensure the best possible thermal contact, these plates were immersed directly in the cooling liquid and the temperature of the sample should therefore have been very close to that of the liquid. The results obtained were quite reproducible even with substances in which the positions of the exciton lines depend markedly on the temperature.

The experiments at the temperatures of liquid hydrogen and liquid helium have been performed in the Aimé-Cotton laboratory at Bellevue and we have pleasure in thanking Professor P. Jacquinot for his interest in our work and for making the facilities of his laboratory available to us.

§ 4. Experimental Observations on Exciton Spectra

4.1. *Preliminary Remarks*

Though exciton spectra were probably observed by Hilsch and Pohl (1928), it is believed that the first clear experimental evidence for line spectra of the exciton type has only recently been presented by Hayashi and Katsuki (1950, 1952), Gross (1956 b), Gross and Karryev (1952 a, b), Gross and Zakharchenya (1953 a, b, 1954) and somewhat later by the present author and his co-workers. In their first paper, Hayashi and Katsuki (1950) published a spectrum of Cu_2O in which exciton lines can be distinctly seen, but they did not attribute them to an exciton spectrum. Line spectra of different substances both in the pure state and after the addition of known impurities are reproduced in the publications of Kokhanenko and Gol'tsev (1952, see also, Kokhanenko 1954). These lines are most probably due to excitons (our raies ultimes) but the above workers suggested that they were due to excess metal. The author is grateful to Professor M. Hayashi and Professor A. V. Terenin for kindly drawing his attention to these earlier observations.

The results obtained in the author's laboratory will be summarized in the following sections. It will be noticed that the Japanese and Russian authors first investigated the spectrum of Cu_2O whereas our work was begun on HgI_2. As the spectrum of this substance proved to be very complicated, it was decided after the publication of the first papers on Cu_2O by Gross and Karryev (1952 a, b), to repeat their observations and then to extend the investigation to other substances. In recent papers Gross and his co-workers have repeated some of our experiments on HgI_2 and PbI_2. They have also extended their own work to CdS and CdI_2. Gross and Kapliansky (1955), Gross, *et al.* (1957), Gross and Rasbirov (1957), see also Broude *et al.* (1957). A part of the work of the Russian group has recently been summarized by Gross (1956 b, see also Apfel and Hadley 1955).

4.2. *The Hydrogen-like Spectra of Cu_2O*

4.2.1. *Experimental observations*

Cu_2O is a substance with which exciton spectra are easily observed Gross (1956 a), Gross *et al.* (1953, 1956), Nikitine (1955 a, b, 1956), Nikitine, Couture, Sieskind and Perny (1954, 1955), Nikitine, Couture, Perny, Sieskind and Reiss (1955), Nikitine, Perny and Sieskind (1954 a, b, c), Nikitine, Reiss and Perny (1955), Nikitine, Sieskind and Perny (1954). Natural and synthetic crystals have been polished to produce sheets with a thickness between 70 and 20 μ. Thin films with a thickness of about 10 μ have been obtained by superficial oxidation of electrolytic copper followed by quenching in liquid air to facilitate the separation of the oxide film. Three hydrogen-like series of absorption lines have been observed. They will be described as the ' red ', ' yellow ' and ' green ' series. The lines

of the first series are very weak and are best observed with rather thick crystals (50–70 μ). Those of the second series are considerably stronger and are observed in moderately thin films (about 35 μ) while the last series of lines is observed only in very thin films (about 15 μ). These series all converge to well defined limits and have a pronounced hydrogen-like character (figs. 2 (*a*, *b*), Pl. 1). At 77·3° and 4·2°K they can be represented by the equations (in cm^{-1}):

	77·3°K	4·2°K
Red series	$\nu_k = 16436 - 475/k^2$	
Yellow series	$\nu_k = 17459 - 795/k^2$	$\nu_k = 17525 - 780/k^2$
Green series	$\nu_k = 18514 - 1252/k^2$	$\nu_k = 18598 - 1242/k^2$

The wavelengths in ångström units of the lines of Cu_2O which have been accurately measured at different temperatures are given in table 1. More lines were observed in all three series but their wavelengths could not be accurately determined.

Table 1

k	Red series		Yellow series		Green series	
	77·3°K	4·2°K	77·3°K	4·2°K	77·3°K	4·2°K
2	6124·5	—	5792	5770·6	5494·2	5468·1
3	6105	—	5756	5735·2	5442	5417·1
4	6095	—	5746	5722·4	5402·5	5401·9
5	6091	—	5737	5716	—	—
6	—	—	—	5713·2	—	—
∞	6087	6057	5727	5707	—	5377(?)

The red series could also be observed at 4·2°K; it was, however, rather weak and the plate has not yet been measured. It is hoped to resume this work with better crystals. The weak lines of these series are usually observed against a background of strong continuous absorption and, for good contrast, it is therefore necessary to determine the optimum exposure for each line. There is another absorption discontinuity in the red, the interpretation of which is uncertain although absorption due to impurity centres might be responsible for it.

The mean difference between the limits of the series, $\Delta\nu_{00}$, is 1050 cm^{-1} and the infra-red frequency, ν_{IR}, of Cu_2O is 1100 cm^{-1}. This suggests that the three series are vibrational satellites. The change in the values of the reduced Rydberg constants R' from one satellite series to another is no argument against this interpretation because satellites in different spectral regions can, in fact, have different values of R'. A considerable temperature variation in the wavelengths of the spectral lines has been found (Nikitine, Sieskind and Perny 1954). Series of lines clearly obeying a formula of the hydrogen type (eqn. (1)) have so far been observed only with Cu_2O.

4.2.2. *Tentative interpretation*

A tentative interpretation of the spectra of Cu_2O is possible in terms of the theory of Elliott (1957). His arguments are plausible and additional supporting evidence is provided by our own unpublished observations.

(1) Though the absolute value of the absorption coefficient of Cu_2O has not been measured either in the lines or in the continuum, it is well known that the absorption is considerably smaller than in other crystals. The green series has to be observed in crystals with a thickness of 15 to 20 μ and the yellow series in crystals with a thickness of 25 to 40 μ. For the red series, still thicker crystals are required. It will become evident that the line spectra in many substances (PbI_2, CuCl, CuBr, CuI, TlCl, TlBr, TlI), are best observed in specimens with a thickness of 0·1 to 0·2 μ, and sometimes even less. The absorption of Cu_2O is therefore several hundred times weaker. The spectra of the above substances probably belong to the first class (§ 2.3.2), in which case the absorption in Cu_2O should be weaker than that of CuCl in the ratio of about $0{\cdot}03\,\kappa^3_{CuCl}/\kappa^5_{Cu_2O}$. The value of κ_{CuCl} is about 10 and values between 7 and 12 have been reported for κ_{Cu_2O} (static dielectric constants). Taking a value of about 9, the above ratio is of the order of 10^{-3} which is roughly in agreement with the observations. It is not certain that the static dielectric constants should be used in both cases so that the importance of this comparison should not be over-emphasized.

(2) Assuming, as we suggest, that the line $\lambda = 6124$ Å belongs to the red series, the first lines of the three series agree very well with Mott's formula with $k = 2$.

(3) The relative intensities of the lines in the series of Cu_2O are not well known. Rough estimates are, however, more in agreement with spectra of the second class than with spectra of the first class in which the first line is much stronger than the succeeding lines.

Elliott proposes to attribute the green and yellow series to splitting due to spin orbit coupling in Cu^{++}. We are of the opinion that this requires further consideration. A somewhat similar attempt has been made by Hayashi and Katsuki (1952) for O^-, but neither interpretation gives an explanation of the red series which is, however, weaker than the other two and has not been studied to the same extent. More information is required before it can be decided whether the existence of this series is to be regarded as a strong argument against Elliott's interpretation. It may be found that this series has to be included in another class of exciton spectra or even excluded as an exciton series. Elliott's theory of spectra of the second class is concerned with transitions with a small change in $\mathbf{k}$. This is, however, not to be compared with the vibrational structure of the spectrum of Cu_2O suggested by the author. The approximation of the theory does not lend itself to the treatment of interactions with optical vibrations, and the author is grateful to Dr. R. J. Elliott for stimulating correspondence on this point.

4.3. *The Dichroism of the Line Spectra of Excitons in Anisotropic Crystals*

4.3.1. *The line spectrum of HgI_2*

The red tetragonal variety of HgI_2 was the first substance which we investigated. The spectrum was found to be dichroic and complicated and it is not yet completely understood (Nikitine 1955 b, Nikitine, Couture, Sieskind and Perny 1954, Nikitine, Couture, Perny, Sieskind and Reiss 1955, Nikitine and Sieskind 1955, Sieskind 1956, Sieskind and Nikitine 1957). The yellow rhombic variety is being studied at present. Fused thin films and thin monocrystals with surfaces both parallel and normal to the optic axis have been prepared. Very thin crystals with surfaces parallel to the optic axis are of particular interest as the ordinary and extraordinary spectra can be observed by turning the analyser through $\frac{1}{2}\pi$. Great difficulties are, however, encountered in the preparation of thin crystals and the spectra are complicated and still need thorough analysis.

When the optic axis lies in the plane of the thin monocrystal, two spectra can be observed at 77·3°K by turning the analyser through $\frac{1}{2}\pi$ to allow the electric vector to vibrate along or normal to the optic axis.

In one of the positions (position 1) of the analyser, a broad line separated from a very strong continuum by a rather weak continuous absorption is observed. The wavelength of this line is $\lambda_1 = 5330$ Å. If the analyser is then turned through $\frac{1}{2}\pi$ (position 3), the line is covered by a strong continuous absorption which has a fairly sharp limit. Very thin crystals are difficult to prepare and only a few have been used for observations at 4·2°K. At 77·3°K, we have examined only rather thick crystals and have not been able to observe the structure of the spectrum in position 3. If the analyser is turned to an intermediate position (position 2), both spectra can be observed simultaneously. It can then be seen that the continuous spectrum observed with the analyser in position 3 has a line at $\lambda = 5367$ Å near the limit of the continuous absorption but shifted slightly to the short wave side of it. Presumably this line at 5367 Å should be observed with the analyser in position 3 if a thin enough crystal of sufficient area could be obtained (see below). It can, in any case, be concluded that both the line and continuous spectra of this substance are dichroic as the spectrum passes from that observed in position 1 to that observed in position 2 when the analyser is turned through an angle of about $\frac{1}{4}\pi$. This situation is illustrated by the diagrams 1, 2 and 3 of fig. 3. In fig. 4, a recording is given of a spectrum obtained in position 2. A third line appears on this curve but it is not quite certain that it is genuine as the corresponding fluctuation in the curve is only about twice as large as fluctuations due to the granularity of the plate.

In order to determine which of the spectra corresponds with the ordinary, and which with the extraordinary ray, two experiments have been carried out. First, a single crystal plate oriented normal to the optic axis was investigated. The spectrum was identical with that obtained in position 3

Fig. 3

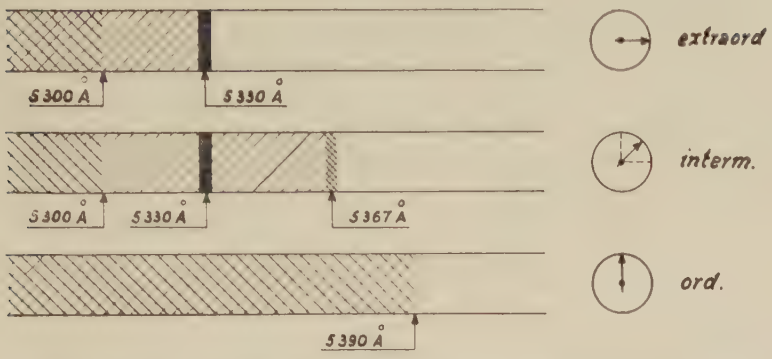
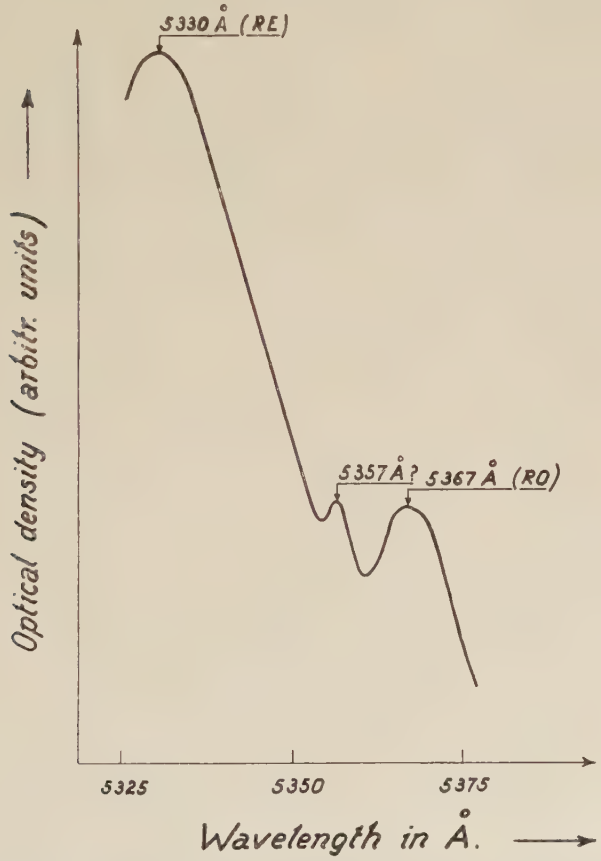


Diagram illustrating the observations with three positions of the analyser with a HgI_2 crystal at 77·3°K.

Fig. 4



Recording of a spectrum of HgI_2 with the analyser in position 2.

of the analyser and no variation with the rotation of the analyser was observed. This spectrum must therefore be attributed to the ordinary ray; the first spectrum containing λ_1 corresponds with the extraordinary ray. The particular single crystal was observed in a polarizing microscope in convergent light and the characteristic interference figure for an uniaxial crystal orientated normal to the optic axis was obtained. These conclusions were confirmed by using a calibrated analyser for the first experiment; the dichroism then appears clearly at 77·3°K.

At very low temperatures, the spectra of HgI_2 are also complicated, but it has been possible to arrive at a number of definite conclusions. Quite recently, Sieskind (1957) has succeeded in growing very thin crystal plates both parallel and normal to the optic axis. Their thickness was measured to a few microns. The crystals normal to the axis are red and their colour does not change when they are observed through a rotating polaroid. The crystals parallel to the axis are rather yellow, when examined by light with the electric vector parallel to the axis, and red when the electric vector is perpendicular to the axis. These crystals were investigated at low temperatures (1·5°K, 4·2°K and 20·4°K). They deteriorated during these measurements and their properties at 77·3°K could not be studied.

The absorption spectrum at 1·5°K of the plates oriented normal to the optic axis has a very strong band at about 5320 Å, a weaker band at 5250 Å, and a band between 4867–4920 Å. In a number of crystals, an absorption step appears at 5230 Å. The first two bands consist of several lines and the fine structure appears distinctly in some crystals. When the incident light is plane polarized, the spectrum does not vary as the plane of polarization is rotated and uniaxial interference figures have been observed with some of the crystals. There is thus no doubt that the crystals are oriented perpendicular to the optic axis, and that the observed spectrum corresponds with the ordinary ray. With these crystals, the second band was considerably weaker than the first. The wavelengths of the lines are reproduced in table 2. The ordinary spectrum has also been observed with crystals oriented parallel to the axis: the lines or bands occur at the same wavelengths as in crystals oriented normal to the axis. In these crystals, which were much thinner than those oriented normal to the axis, the first band was much weaker than the second. A similar effect has been observed with other substances.

For these very thin crystals, the extraordinary spectrum consists of a rather weak line at about 5298 Å, near which a second very weak line is sometimes observed. An absorption step occurs at about 5050 Å. The extraordinary spectrum was observed at an earlier stage of the work with considerably thicker crystals (Nikitine and Sieskind 1955). At 4·2°K, the wavelength of the line was 5296 Å and it was accompanied by a weaker line 5309 Å. A third extremely weak line was observed at 5321 Å, but it is likely that this line belongs to the ordinary spectrum which is recorded in table 2 and appears because a convergent beam of light was used for

this experiment. With these thick crystals, the ordinary spectrum could not be clearly observed as the continuous absorption between the bands was too strong. The wavelengths of the lines, according to our recent measurements, are reproduced in table 3. There is a discrepancy of about 2 Å between these values and our earlier measurements.

The different absorption spectra of HgI_2 are illustrated in figs. 5 and 6, Pls. 1 and 2 and fig. 7. The temperature shift of the lines has been investigated by Sieskind and will be reported elsewhere.

Table 2. Ordinary Spectrum (Wavelengths in Å)

	20·4°K	4·2°K	1·5°K	Mean wavelength of the bands
1st band	5337 5324 5320 5314	5322 5319 5316	5323 5320 5316	5320
2nd band	5258 5244	5253 5248	5253 5248 abs. step 5230	5250
3rd band	from 4867 to 4920	from 4870 to 4970	no clear observations	

Table 3. Extraordinary Spectrum (Wavelengths in Å)

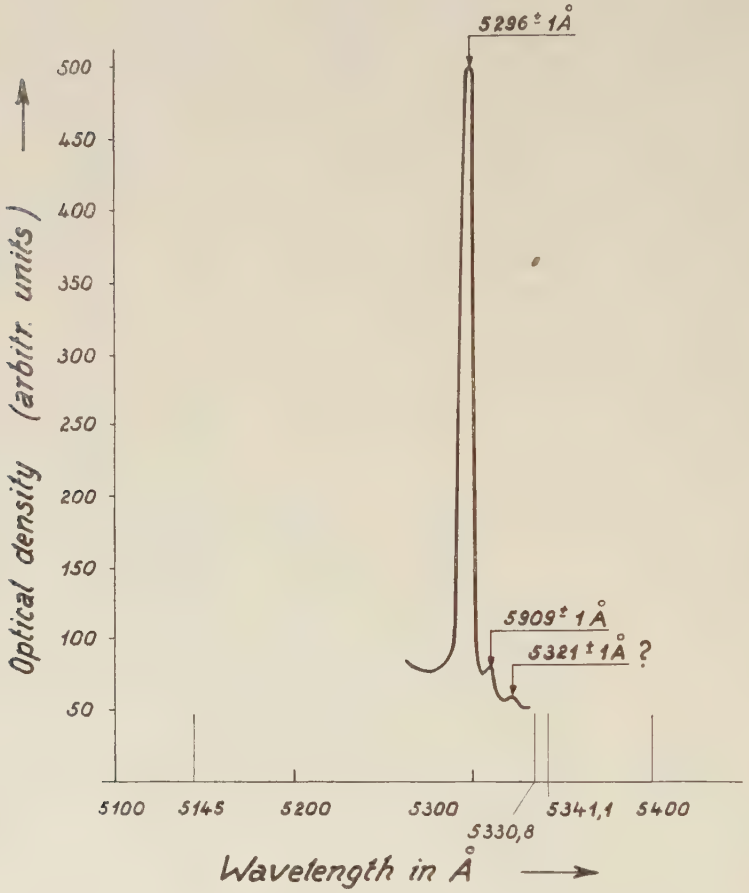
	20·4°K	4·2°K	1·5°K
Strong line	5301	5298	5300
Very weak line	—	5312	5313
Limit of continuous absorption about	5050	5050	5050

Gross and Kapliansky (1955) have repeated our experiments on HgI_2 and have not confirmed some of our results. According to them, the line 5298 Å belongs to the ordinary spectrum and the group of lines at 5320 Å does not exist. Our recent observations are, however, in good agreement with those presented in our earlier publications and we have no doubt that the two spectra should be attributed to the ordinary and to the extraordinary absorption. We have also no doubt over the existence of the group of lines at 5320 Å as seen in fig. 5.

In our recent investigations, spectroscopically pure material from Johnson Matthey has been used in conjunction with very pure solvents from Merk. No significant difference has been found between these and

the earlier observations with less pure crystals. The above results are in good agreement with the measurements of the reflection spectra which will be reported elsewhere. We shall give a tentative interpretation of the spectra of HgI_2 in § 4.3.3.

Fig. 7



Recording of the group of lines reproduced in fig. 6.

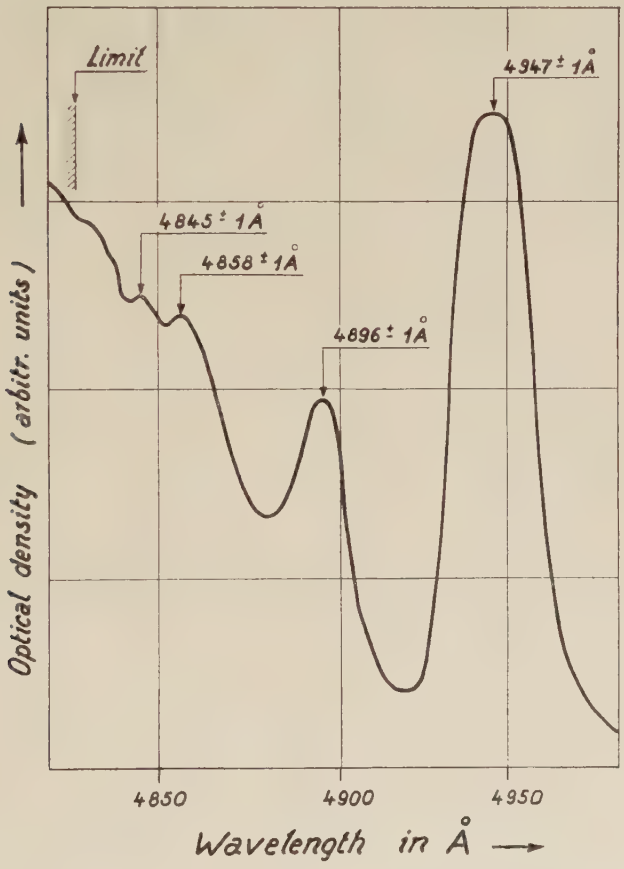
4.3.2. *Line spectra of* PbI_2

This substance crystallizes in the hexagonal system and it is possible to obtain thin sheets by fusion, by condensation from the vapour or by crystallization from solution. The single crystals grown from solution are all oriented normal to the optic axis and their strong absorption corresponds to the ordinary spectrum. In plates with a thickness of about 2 μ, we have observed a strong absorption line at the limit of a strong continuum (4947 Å at 4·2°K). With thin microcrystalline sheets, the extraordinary and the ordinary spectra are superposed. A strong isolated line, identical with that observed in the ordinary spectrum of single

crystals, and a hydrogen-like series of lines have been observed under these conditions. These lines occur at the following wavelengths: $\lambda_2 = 4896$ Å; $\lambda_3 = 4858$ Å; $\lambda_5 = 4845$ Å $\pm$ 1 Å and there may be two more absorption lines. In some experiments, another line $\lambda_x = 4917$ Å was observed between the strong line and the first line of the series. The spectrum of the PbI_2 may also be dichroic, but no clear evidence for this has been obtained as crystals orientated parallel to the optic axis have not yet been prepared. The series is represented by the equation:

$$\nu_k = 20711 - \frac{1144}{k^2}\ \mathrm{cm}^{-1}, \qquad k = 2, 3, 4, 5 \ldots\ldots$$

Fig. 9



Recording of the first spectrum reproduced in fig. 8.

The line $\lambda_1 = 4947$ Å does not correspond to this equation but Haken has shown that good agreement would not be expected. This may therefore be the line corresponding to $k = 1$, of the above series, displaced towards a shorter wavelength from the predicted position. The same first line

$\lambda_1 = 4947$ Å, is observed with very thin sheets prepared by evaporation. In such sheets, the continuum becomes very weak and the absorption in this part of the spectrum is due effectively to this narrow line. Further details will be found in the original papers (Nikitine 1955 b, 1956, Nikitine, Couture, Perny, Sieskind and Reiss 1955, Nikitine and Perry 1955 a, b, c, 1958.) Both kinds of spectrum are reproduced in fig. 8, Pl. 2 and a recording of the first spectrum in fig. 9.

4.3.3. *Tentative interpretation of the exciton spectra of HgI_2 and PbI_2*

It can be seen that the characteristics of the spectra of both HgI_2 and PbI_2 are those of the first class. The absorption is very strong and the first lines of the series are clearly visible even in crystals with a thickness of less than 1 μ. The intensity of these lines is probably at least an order of magnitude greater than that of the other lines of the series.

With PbI_2, the second line corresponds to $k=2$ in Mott's equation, so that the first line could be the line corresponding to $k=1$ shifted from its theoretical position. The lines of the ordinary spectrum of HgI_2 can be represented by a series formula. Neglecting the fine structure and using the wavelength of the first line in each multiplet, it can be seen that the lines 5322 Å, 5253 Å and the step at 5230 Å are given by the equation:

$$\nu_k = 19{\cdot}119 - 330/k^2, \qquad k = 1, 2.$$

The lines $k=3, 4$ are not observed, possibly because they are too weak. This suggests that the spectrum belongs to the first class but no simple explanation can be given for the change in the relative intensities of both groups of lines in the thin crystals which is described above.

Only one group of lines has been observed in the extraordinary spectrum of HgI_2 and no classification is possible. If these are the first lines of a series ($k=1$), the other lines may be too weak for observation against the continuous background absorption. No theoretical interpretation can be given for the fine structure of the lines, although a treatment similar to that of the fine structure of atomic spectra can be envisaged. A theory of the fine structure of exciton lines has, however, not yet been developed. Mott's theory has recently been extended to anisotropic media and qualitative agreement with the observations can be reported (Barriol *et al.* 1956). The theory is not directly applicable to iodides so that a direct comparison has not been possible.

4.4. *The Spectra of Excitons in Isotropic Crystals*

4.4.1. *Exciton spectra of CuI*

CuI has the zincblende structure. Good films can be prepared either by melting the material between Pyrex plates or by condensing it from the vapour on to a glass plate at a temperature of about 120°C. Films prepared by the first method are usually rather thick and are probably strained but very thin films can be easily prepared by the second method. These give well-developed systems of interference fringes and their thicknesses can

be evaluated, sometimes with high accuracy. When some of these films were annealed, recrystallization was observed with the naked eye; the superficial reflection became diffuse, the fringes decreased in visibility and they sometimes even disappeared completely. The specimens have been prepared from pure CuI and from spectroscopically pure CuI from Johnson Matthey; no systematic differences have been observed with substances of different purity.

No final conclusions can be drawn about certain aspects of the spectral properties of CuI as they are influenced by the methods used for the preparation of the films; the general features, however, will be apparent from the results which are presented below (Nikitine 1956, Nikitine, Couture, Perny and Reiss 1955, Nikitine, Couture, Perny, Sieskind and Reiss 1955, Nikitine and Reiss 1956 a, b, 1957 a, b, c, Nikitine, Reiss and Perny 1956 a, b, Reiss 1956). Films prepared by the different methods have complicated line spectra at 4·2°K. The lines are very narrow ($\Delta\lambda_{1/2} \simeq 1$ Å), and their wavelengths are given in table 4. The f samples were fused films; the e samples were obtained by evaporation; a represents annealed, c, continuum, and g, gap in the continuous absorption. The sample 1 f was thinner than the other two f samples. 1 e was rather thick; 2 e was annealed and about 2 μ in thickness; 3 e was about 0·4 μ in thickness and 4 e less than 0·1 μ. With the thicker films, the lines appear near the limit of a branch of the continuous absorption and the exposure has to be determined for each individual line to ensure optimum contrast on the photographic plate.

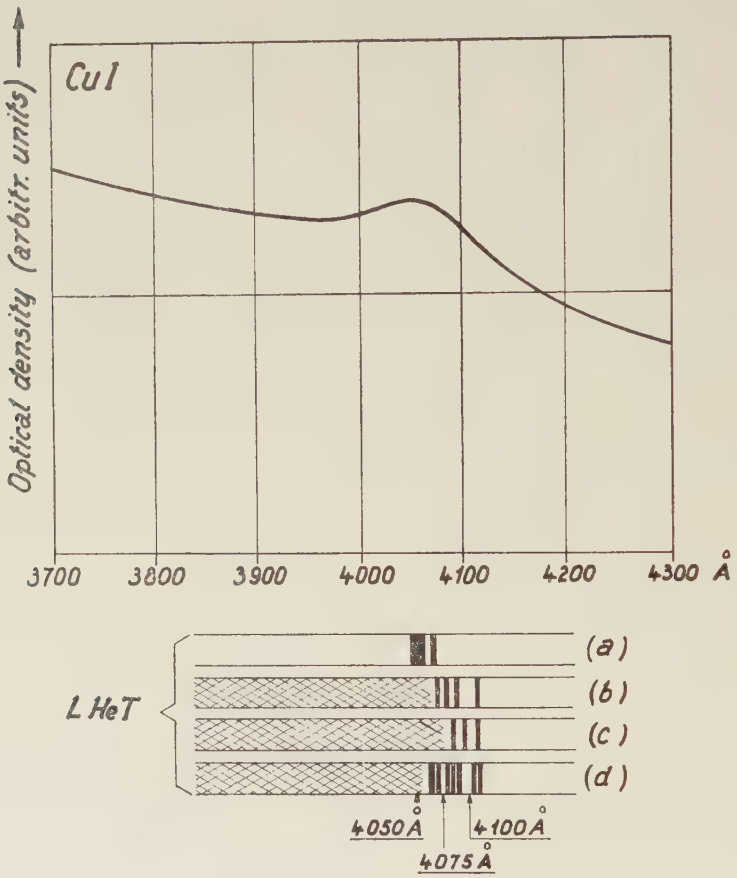
Table 4

Specimen							
1 f	4106, 4102		4083, 4079		4078 ?	4069, 4064 c	g 4031 c
	$\Delta\lambda$=4 Å		$\Delta\lambda$=4 Å			$\Delta\lambda$=5 Å	
2 f	4106	4093 c from		4080			
3 f	4106	4093	4083 c				
1 e	4106	4093		4079	c from 4079		g 4032 c
2 e a	4106		4083	4075		4065·6 c	g 4032 c
3 e						4065·6	4051 no c
4 e							4051 no c

The doublet structure which appears distinctly with specimen 1 f does not appear with the other specimens. The lines 4106, 4065·6 (or 4064) Å and the line 4051 Å, which is usually broad, appear with all the samples provided that they are of a suitable thickness. The lines 4102 and 4069 Å appear only in sample 1 f. The other lines do not always appear; they are nevertheless sharp and have a well-defined wavelength. We shall refer to them as structure sensitive or irregular lines. The photographs of the spectra of specimens 1 e and 2 e are reproduced in fig. 10, Pl. 2. In fig. 11, a schematic representation of the spectrum of CuI and part of the curve given in the paper by Fesefeldt and Gyulai (1929) are reproduced.

In recent experiments (Nikitine and Reiss 1957 a, b, c, d), the absorption has been determined as a function of the thickness of the specimen at 77·3°K. At this temperature, the majority of the lines of CuI are observed only with difficulty, but the two lines at 4063 Å and 4051 Å (there is no considerable temperature shift) are clearly recorded when the specimen

Fig. 11



Schematic representation of the exciton spectra observed in CuI at 4·2°K compared with the continuous curve obtained by Fesefeldt and Gyulai (1929) ; though completely different this curve shows a slight inflection in the position in the spectrum corresponding with our observations.

has a suitable thickness. A continuous spectrum, which extends to longer wavelengths, obscures these two lines when the thickness exceeds several microns although they can be observed in the form of a single broad band up to thicknesses of one or two microns. The continuous spectrum begins on the short wavelength side very close to the lines but becomes

weaker as the thickness decreases. The band then breaks up into two components. A second doublet at 4002 Å and 3990 Å is observed in specimens with a thickness of about 0·5 μ. At 4·2°K, the continuum then begins about 3970 Å. Continuous absorption has also been observed at 77·3°K. When the thickness of the specimen is further decreased, the continuum becomes very weak and is scarcely in evidence in specimens with a thickness between 0·1 and 0·005 μ. In this range of thickness, another strong line appears at 3341·5 Å. Figures 12 (*A*), (*B*), Pl. 3, and 13, Pl. 4, illustrate the observations.

It seems probable that these lines belong to a class which is different from that of the other lines of the spectrum of CuI. They are quite reproducible and we have suggested that they should be described as 'raies ultimes'. The value of the oscillator strength, $f \simeq 2 \times 10^{-3}$, for the doublet 4063 and 4051 Å has been calculated from anomalous dispersion curves deduced from interference fringes and from anomalous reflection (Nikitine and Reiss 1956 a)†. If the light reflected from the surface of a crystal of CuI is analysed with a spectrograph, a strong maximum of reflection, analogous to the 'residual rays', is observed at approximately the same wavelength as the absorption line observed in transmission. A very sharp minimum of reflection is observed on the short wavelength side of the absorption line and it has been suggested that such minima should be described as 'missing rays' (rayons manquants). The phenomena, which can be interpreted on the basis of anomalous dispersion (Nikitine and Reiss 1956 a), have been observed with a number of substances when the absorption lines are very strong and belong to the class of 'raies ultimes'. No observations of anomalous reflection have been made in the vicinity of the structure sensitive or irregular lines. The spectrum of CuI is thus quite complicated and it is evident that the structure sensitive or irregular lines and the 'raies ultimes' belong to different classes.

No simple interpretation of these spectra can be given. Two possible interpretations may be suggested depending upon whether or not Lambert's law is applicable to thin crystals. The validity of this law has not been established for the thinnest crystals which have been examined.

(1) If we suppose that Lambert's law applies over the whole range of thicknesses used in our experiments, the results obtained with the thinnest specimens can be extended to all the specimens independently of their thicknesses. The first 'raies ultimes', 4063 and 4051 Å are then 50 to 100 times stronger than the continuum and the structure sensitive or irregular lines. The f value for these lines is of about 2×10^{-3}. The first doublet at 4063 and 4051 Å, the second doublet at 4002 and 3990 Å and the continuum at 3970 Å can be represented by the equations:

$$\nu_k{}^1 = 25{\cdot}189 - \frac{504}{k^2}\ \mathrm{cm}^{-1}, \qquad k = 1, 2$$

and

$$\nu_k{}^2 = \nu_k{}^1 - 78\ \mathrm{cm}^{-1}$$

† *Note added in proof.*—A value $f = 6 \times 10^{-3}$ has been obtained from recent and more accurate measurements at 77·3°K.

which we suppose can be extended to the line $k=1$, despite the remarks of Haken. The doublets corresponding to $k=3$, 4 are not observed. They may be too weak or else appear only at very low temperatures. The ultra-violet line probably belongs to another spectrum of the same class. The difference in energy between the first doublet and the ultra-violet 'raie ultime', 0·64 ev, can be compared with the difference between the two spectroscopic states of the neutral iodine atom which is 0·9 ev. All these properties agree very reasonably with what would be predicted by the theory for exciton lines of the first class, and we believe that this is the simplest interpretation of the raies ultimes. If the interpretation of the 'raies ultimes' of CuI is correct, the remaining much weaker structure sensitive or irregular lines probably have a different origin.

(*a*) They are not likely to belong to the second class of transitions discussed by Elliott but might be included in another class of exciton lines arising from forbidden transitions, occurring at defects and dislocations.

(*b*) Cuprous iodide can also contain an excess of iodine, which could introduce acceptor centres just above the valence band of the crystal. These would be unoccupied at low temperatures and the structure sensitive or irregular lines might be due to electronic transitions from the valence band. It is hoped that further information on the structure sensitive lines will be obtained from the experimental work which is in hand to test these hypotheses.

(2) Although none of the experimental results is inconsistent with the above interpretation, there is an alternative interpretation which should be considered. As has already been mentioned, it is not certain that Lambert's law can be applied to specimens with very small thicknesses and the validity of the law is at present being investigated experimentally. Until the results are available, there must be some doubt as to the validity of the law for very thin specimens. The diameter of an exciton orbit is $D=\kappa k^2$ Å; with $\kappa=10$, D will be of the order of several 100 A. If excitons are formed consequently only in the interior of a thin sample, the effective volume will be a small fraction of the total volume. For these reasons, one might anticipate that, for the thinnest specimens used in our experiments, Lambert's law would not apply to the exciton lines which correspond with high values of k. If moreover, there are observable optical transitions to surface states, it is clear that their importance, relative to exciton transitions occurring within the crystals, will increase with decreasing thickness of the specimen. Our samples have been obtained by condensation from the vapour and may be somewhat porous. The greater surface area of a porous film compared with a compact film of the same nominal thickness will result in a relative enhancement of the lines corresponding to surface transitions if these are important. This is consistent with the above arguments.

If these considerations are relevant, one may interpret the 'raies ultimes' as transitions involving surface states. In thin films, these

lines will become relatively more important than the fundamental absorption in the volume of the crystal, whereas their importance in thick films will be small in comparison with the fundamental absorption. This last point can, unfortunately, probably not be studied experimentally as the absorption is too great in thick films. According to this interpretation, the structure sensitive or irregular lines must be attributed to the fundamental absorption and assigned to a class of exciton lines not already covered by Elliott's theory.

It is difficult to decide in favour of one or the other suggested interpretation. More experimental results are required but, in the present state of our knowledge, the first interpretation is considered to be more likely for the following reasons:

(*a*) Very little is known of the surface states and of the optical transitions to such states which are introduced in the second interpretation.

(*b*) The 'raies ultimes' are satisfactorily interpreted as exciton lines of first class.

These considerations support the first interpretation; the alternative interpretation should, however, be borne in mind, because of the doubt which exists as to the validity of Lambert's law for very thin specimens. The lack of reproducibility of some of the weaker lines probably depends on effects due to crystal imperfections, the importance of which is not yet known. If this property should be established, the occurrence and the intensities of these lines may provide a very sensitive measure of the degree of imperfection of the crystals. It will be noticed that similar irregular lines have been observed by Gross (1956 b) and his co-workers and in the author's laboratory for Cu_2O and CuBr crystals, and by different Russian groups for CdS crystals (Gross and Kapliansky 1955, Gross *et al.* 1957, Gross and Rasbirov 1957, Broude *et al.* 1957). These lines are thus not accidental, but represent a class of absorption phenomena occurring in a number of substances. Many lines of this type are observed in the spectra of specimens of suitable thickness. They form a series which converges to a continuum covering the 'raies ultimes'. On the basis of our first experiments we have made a tentative classification of the lines into two hydrogen-like series which in fact form doublets. In the thin film of cuprous iodide, 2 e, the four terms of the series are reproduced. The fourth line, which lies near the limit of the series, does not agree with a hydrogen-like formula. As there is a continuum on the short wave side of this line, it is doubtful whether the proposed classification can be maintained for higher values of the quantum number. No pronounced shift of the spectrum has been observed between 77·3 and 4·2°K.

4.4.2. *Investigation of exciton spectra of CuBr and CuCl*

The spectra of CuBr and CuCl are being studied in the author's laboratory at present but only preliminary results have so far been obtained (Nikitine and Reiss 1956 b, 1957 d). These salts are hygroscopic and difficulties are

encountered in the experimental work. Thin films of both substances were prepared by evaporation in vacuum on to heated quartz plates. Interference fringes were observed and the thicknesses of the films could be measured with some accuracy. Fused films of CuBr have also been prepared.

The following lines have been observed at 4·2°K with specimens of CuBr having a thickness between 0·1 and 0·4 μ. The first group is formed by at least three very strong lines:

$$\lambda_3 = 4170\,\text{Å}, \quad \lambda_4 = 4163\,\text{Å} \text{ and } \lambda_5 = 4158\,\text{Å}.$$

Two weak lines are observed at $\lambda_6 = 4129$ Å and $\lambda_7 = 4059$ Å, together with an absorption band at the limit of the ultra-violet, between $3977\,\text{Å} > \lambda_8 > 3960\,\text{Å}$. A rather weak continuous absorption occurs at shorter wavelengths. With a fused specimen a few microns in thickness, two weak and narrow lines have been observed at 20·4°K with wavelengths $\lambda_1 = 4251$ Å and $\lambda_2 = 4225$ Å, together with a continuum beginning at about 4200 Å. In a thin part of this specimen, two lines have been observed at 4190 Å and 4171 Å, these lines probably corresponding to two of the group λ_3, λ_4, λ_5, shifted as a result of the change of temperature. At 77·3°K with yet another specimen, a strong, broad line has been observed between 4170 Å and 4116 Å (to be compared with 4129 Å). It is not possible to study the same specimen at different temperatures as the film usually deteriorates during an experiment. Anomalous reflection is observed in the vicinity of strong lines.

The spectrum of CuBr is similar to that of CuI. There is an obvious analogy between the strong lines of CuBr and the 'raies ultimes' of CuI. It is not yet clear that the weak lines of the spectrum of CuBr have the same properties as the 'structure sensitive or irregular lines' of CuI, and further observations are needed. It is, in any case, evident that the lines belong to different classes; the strong lines probably belong to the first class but the weak lines do not belong to the second class. In the present state of our knowledge it would not be profitable to attempt a more detailed interpretation.

The CuCl films prepared by evaporation had a thickness of about 0·1 μ. The spectrum is similar to that of CuI films produced in the same way. Two strong lines have been observed, the first about 4 Å broad with $\lambda_1 = 3849$ Å and the second about 30 Å broad with $\lambda_2 = 3765$ Å. No continuum has been observed in thin films. The reflection spectrum is also characteristic and shows two pronounced maxima and minima corresponding to residual and missing rays. This effect will be described later. The spectra of thin films of CuI, CuBr and CuCl produced by condensation of the vapour all have a similar structure which is probably characteristic of films of copper halides prepared by this technique.

The spectra at 4·2°K of two thin parts of a specimen of CuBr, are reproduced in fig. 14, Pl. 4. One part was thinner than the other. An absorption and a reflection spectrum of a very thin film of CuCl are reproduced in fig. 15, Pl. 4.

4.4.3. *Line spectra in thallium halides*

TlI, TlBr and TlCl are being investigated at present (Nikitine and Reiss 1956 b). With a fused film of TlI around 1 μ thick, a pronounced line at $\lambda = 4305$ Å, about 30 Å broad, is observed at 77·3 °K; it is shifted to $\lambda = 4291$ Å at 4·2°K. The line disappears, however, in much thinner films in which other lines have been observed (at $\lambda = 4080$ Å). Fused films of TlBr and TlCl of about the same thickness, showed no line structure in the absorption spectrum. At 77·3°K, in films of much smaller thickness (about 0·1 μ) a broad and diffuse line is observed at $\lambda = 4068$ Å for TlBr and at $\lambda = 3597$ Å for TlCl. The line observed in TlI corresponds to a very slight inflection in the absorption curve obtained by Fesefeldt (1930, 1931). The lines found in TlBr and TlCl have not previously been reported. Some of these spectra are illustrated in fig. 16, Pl. 5.

4.4.4. *Line spectra of AgI*

AgI is being investigated by Perny (1958, see also, Perny and Nikitine 1957). Thin films (0·1 μ to 1 μ) have been prepared by different techniques. Under well defined conditions, Perny has been able to prepare micro-crystalline films of $\alpha-$, $\beta-$ and $\gamma-$ AgI and has confirmed the structures by electron diffraction. These structures appear to persist at low temperatures. At these temperatures Perny has observed three completely different spectra each corresponding to the films prepared in the above conditions. He has tentatively concluded that the three spectra in question could be attributed to these three structures respectively, remaining at least in part at low temperatures. At 4·2°K, the following spectra are observed for γ-AgI; two lines at $\lambda_1 = 4270$ Å and $\lambda_2 = 4262$ Å, at the edge of an absorption band, the maximum of which is at $\lambda = 4212$ Å. On the short wavelength side of this absorption band, a convergent hydrogen-like series of lines is observed: $\lambda = 4187$ Å, 4172 Å, 4162 Å, 4154 Å and 4151 Å, the last wavelength probably corresponding to the limit of the series. These lines can be represented by a hydrogen-like series,

$$\nu_k = 24090 - 1790/k^2, \qquad k = 3, 4, 5, 6.$$

Neither the line $k = 2$, which may overlap with the absorption on the long wave side of the series, nor the line $k = 1$ could be observed. There is practically no continuum on the short wavelength side of this series in thin films but another continuum appears at about 4050 Å and extends to the ultra-violet. The experimental data agree with the formula to within about 10 cm^{-1}. For α-AgI, a continuous spectrum is observed from $\lambda = 5005$ Å to the ultra-violet. The absorption spectrum which we attribute to β-AgI is also a continuum extending from $\lambda = 4367$ Å to the ultra-violet. The wavelengths of these lines and bands are dependent upon the temperature. The spectrum of γ-AgI is reproduced in fig. 17, Pl. 5.

4.5. *Luminescence Spectra and Exciton Absorption*

During the researches on exciton absorption, luminescence has recently been observed with the above substances in different laboratories

(Arkhanguelskaya and Feofilov 1956, Grillot 1956, Nikitine 1956, Nikitine and Perny 1958, Nikitine and Reiss 1956 c, Reiss 1956, Sieskind 1957). The luminescence spectra are complicated; in many cases, emission lines (α-lines) are observed in the immediate vicinity of exciton lines with a small Stokes shift. Broader lines (β-lines), apparently of another type, appear in a spectral region where no absorption is observed. The connection between the α-lines and the exciton spectra is obvious; they correspond to the decay of the excitons by radiation. The interpretation of the β-lines is not evident although lines emitted by trapped excitons which had polarized the surrounding crystal might be envisaged. These bands could, on the other hand, be due to luminescence centres within the crystal as all these spectra depend on the technique employed for the preparation of the specimen. The study of the relation between luminescence spectra and the absorption spectra of excitons is very promising, even though the complications of the spectra may be great and their detailed interpretation laborious.

4.6. *Remark on the Purity of the Specimens*

For our preliminary experiments, we used pure materials which had, however, not been spectroscopically standardized. In our recent experiments, we have used spectroscopically pure CuI, HgI_2, PbI_2 supplied by Johnson Matthey & Co. Ltd. The analytical report of the firm states that no impurity could be detected in the above substances in a quantity greater than one part in 10^6. For the preparation of single crystals of HgI_2 from solution, a very pure solvent supplied by Merk was used. The purity of this solvent was, however, not quite as great, as water was present to the extent of about one part in 10^4. In all these cases, it is believed that the impurity content was not increased during the preparation of the specimens.

Specimens have been made from films of electrolytic copper by reaction with iodine vapour or oxygen, and by the simultaneous evaporation of copper and iodine. Fused films have also been prepared from pure and spectroscopically pure substances. It is always difficult to know whether a given specimen has the stoichiometric composition. There was never any observable difference between specimens prepared from spectroscopically pure and from somewhat less pure substances either with the 'raies ultimes' or with the weaker lines and the structure sensitive or irregular lines. The differences which were sometimes observed probably arose from variations in the perfection of the crystals rather than from impurities in the materials from which they were prepared.

We therefore conclude that our results are independent of the residual impurity content of the specimens and that the absorption spectra arise from the fundamental absorption of the pure substances. We cannot be as definite with regard to the stoichiometric composition of the specimens but some specimens were annealed in vacuum and in halogen vapour without any significant modification of the absorption spectrum. Perny

(1958) has, however, reported several new anomalies in the reflection spectrum of a AgI specimen annealed in iodine vapour. This point is of great interest and is still under investigation. Particular difficulties arise with Cu_2O; the stoichiometric ratio is difficult to establish so that the composition may depend upon the technique used for the preparation of the specimens. Experiments are being carried out on the preparation of Cu_2O under well defined conditions. Specimens prepared by the different techniques always give the same green and yellow spectra, but this may not be true of the red spectrum which has not yet been extensively studied.

The majority of our results have therefore probably not been influenced by the presence of impurities even in the case of Cu_2O. Some of the weak lines may depend upon deviations from the stoichiometric composition but there is no definite evidence for such an effect. The state of crystalline perfection in the specimens appears to influence the results to a greater extent than the impurity content.

§ 5. Conclusions and General Remarks on the Spectra of Crystals

Even though the study of the absorption spectra of crystals is a new subject, it is now possible to draw a number of conclusions and to make some general observations on the available results. The absorption coefficient for most of the crystals which have been investigated has a value of the order of 10^5 or 10^6 cm^{-1}. The absorption curve can therefore not usually be measured with thick crystals, the only exception to this being provided by Cu_2O which has a much weaker absorption. The measurements can, of course, be made with thin crystals but Lambert's law may not apply in the range of thicknesses down to 0·1 μ or less which are used for such measurements. Reasons for this are given above and measurements are being carried out in the author's laboratory to establish the range of validity of the law.

In moderately thin crystals, the absorption spectrum at very low temperatures usually consists of a continuum more or less sharply defined at the long wavelength limit together with a line spectrum located on the long wave side of this limit. There is usually a more or less strong background absorption between the lines which are sometimes very narrow (a few Å). They often form series which converge to the limit of the continuous absorption, but these are of the hydrogen-like type only in exceptional cases.

The line spectra which have been observed can be grouped into three classes:

(1) When the spectra are studied with specimens of different thicknesses, it is observed that, for one class of substances, the continuous absorption is weak in thin specimens and can scarcely be observed for thicknesses of 0·1 μ or less. With these substances, strong absorption lines or multiplets

are still observed which have an f value of about 10^{-3} and we have suggested that they should be called 'raies ultimes'. We believe that they could be used for analytical purposes as they are characteristic of the different substances. With thicker specimens, more lines are sometimes observed on the short wavelength side of the 'raie ultime'. They may well form a series of which the 'raie ultime' is the first line (or group of lines). These series have usually no pronounced hydrogenic character although the first line may correspond to $k = 1$ in Mott's formula. Assuming the validity of Lambert's law for all thicknesses, it can be concluded that the first line of the series is much stronger than either the other lines or the continuum. These spectra of the first class are observed with a great number of substances and particularly with halides.

(2) The second class is represented by the spectrum of Cu_2O. In this substance, three series of lines having a pronounced hydrogen-like character are observed. The lines are all of comparable intensity, of the same order as that of the continuum. The first line in each series corresponds to $k = 2$ in Mott's formula. The absorption coefficients are about 100 times weaker than for the spectra of first class.

(3) In some substances which have a line spectrum of the first class, additional weaker lines are sometimes observed with specimens having a thickness about 10 times greater than that suitable for the observation of 'raies ultimes' (CuI, CuBr, CdS). These lines, which are usually numerous, are probably more than 10 times weaker than the 'raies ultimes' and we suggest that they should be grouped in a third class. They have comparable intensities and form converging series which are not necessarily of hydrogenic character. Their wavelengths are well defined and quite reproducible even though their intensities are not. The factor which governs the intensities is not yet known but it may be the state of imperfection of the crystal determined by the number of dislocations or defects per unit volume. We propose to call these lines 'structure sensitive' or 'irregular' lines.

A theoretical interpretation of these spectra can be attempted on the basis of the recent theories of transition probabilities due to Elliott and Haken. The first and second class of line spectra are accounted for by the first and second classes of exciton spectra in Elliott's theory and these spectra are therefore probably correctly interpreted as exciton spectra. This interpretation of the spectra of first class is, however, correct only if Lambert's law is valid for all thicknesses. If Lambert's law does not apply, an alternative interpretation of the 'raies ultimes' as transitions to surface states has been suggested. At present, the first interpretation seems more reasonable as there is good agreement with Elliott's theory and surface states about which little is known are not introduced. If this interpretation is correct, the 'raie ultime' is the first line ($k = 1$) of a series of exciton lines of the first class. The value of the oscillator strength deduced from anomalous reflection and dispersion for the first line of

CuI is also in good agreement with Elliott's theory but the splitting of the 'raies ultimes' into multiplets is not yet explained.

The characteristic properties of the spectra of second class correspond to those predicted by Elliott's theory for the formation spectra of excitons of the second class. The reason why three equidistant series are observed in Cu_2O is, however, not yet clear. The interpretation of the lines of the third class cannot be based on any theory which has so far been advanced. They certainly do not belong to the exciton spectra of first or second classes of Elliott's theory but may represent a third class of exciton lines normally corresponding to forbidden transitions which occur under the perturbing influence of dislocations or lattice defects in crystals. Such transitions are not covered by Elliott's theory. These lines could also represent transitions from the valence band to acceptor centres and more information is required before a satisfactory interpretation can be given.

Reflection spectra have also been studied in the author's laboratory. It can be shown that anomalous reflection (residual rays and missing rays) is usually observed in the vicinity of strong lines such as the ' raies ultimes '. It is possible that these reflection spectra can also be used for analytical purposes. This aspect of the problem will be described in another publication. Luminescence spectra, obviously related to exciton spectra, have also been observed; they are complicated and will be discussed elsewhere.

As many experiments have been carried out with spectroscopically pure substances and great care taken in preparing the specimens, the above results are unlikely to have been affected by impurities. Deviations from the stoichiometric composition and crystal imperfections probably have some influence on the spectra though there is no definite evidence for this at present. Spectra have been obtained with thin films produced by condensing the vapour on substrates, and these may not have the compact structure of ordinary crystals. This may also be important, but no systematic studies which would establish the existence of such effects have yet been undertaken.

The absorption spectra of many ionic crystals have now been measured. Further experimental and theoretical work will contribute to a better understanding of the spectra but the author is of the opinion that the interpretations which have been proposed are reasonable and provide convincing evidence for the existence of exiton spectra.

Acknowledgments

The author is greatly indebted to Professor N. F. Mott and Dr. J. W. Mitchell for the revision of the English manuscript and for very useful discussions and to Dr. R. J. Elliott and Dr. H. Haken for the communication of theoretical papers in advance of publication as well as for stimulating discussions and important private communications.

References

Apfel, J. H., and Hadley, L. N., 1955, *Phys. Rev.*, **100,** 1689.

Arkhanguelskaya, V. A., and Feofilov, P. P., 1956, *J. Phys. Radium*, **17,** 824.

Barriol, J., Nikitine, S., and Sieskind, M., 1956, *C.R. Acad. Sci., Paris*, **242,** 790.

Broude, V., Eremenko, V., and Rashba, E., 1957, *Dokl. Akad. Nauk S.S.S.R.*, **114,** 520.

Dexter, D. L., 1951, *Phys. Rev.*, **83,** 435.

Dresselhaus, G., 1956, *J. Phys. Chem. Solids*, **1,** 14 ; 1957, *Phys. Rev.*, **106,** 76.

Dykman, I. M., and Pekar, S. I., 1952, *Dokl. Akad. Nauk S.S.S.R.*, **83,** 825.

Elliott, R. J., 1957, *Phys. Rev.*, **108,** 1384.

Fesefeldt, H., 1930, *Z. Phys.*, **64,** 741 ; 1931, *Ibid.*, **67,** 37.

Fesefeldt, H., and Gyulai, Z., 1929, *Nachr. Akad. Wiss. Göttingen*, 226.

Frenkel, J., 1931 a, *Phys. Rev.*, **37,** 17 ; 1931 b, *Ibid.*, **37,** 1276.

Grillot, E., 1956, *J. Phys. Radium*, **17,** 822.

Gross, E. F., 1956 a, *Izv. Akad. Nauk S.S.S.R., Ser. fiz.*, **20,** 89 ; 1956 b, *Suppl. Nuovo Cimento*, **3,** 672.

Gross, E. F., and Kapliansky, A. A., 1955, *J. tech. Phys., Moscow*, **25,** 2061.

Gross, E. F., and Karryev, N. A., 1952 a, *Dokl. Akad. Nauk S.S.S.R.*, **84,** 261 ; 1952 b, *Ibid.*, **84,** 471.

Gross, E. F., Rasbirov, B. S., and Jacobsen, A., 1957, *J. tech. Phys., Moscow*, **27,** 1149.

Gross, E. F., and Rasbirov, B. S., 1957, *J. tech. Phys., Moscow.*, **27,** 2175.

Gross, E. F., and Zakharchenya, B. P., 1953 a, *Dokl. Akad. Nauk S.S.S.R.*, **90,** 745 ; 1953 b, *Ibid.*, **92,** 265 ; 1954, *Ibid.*, **99,** 231.

Gross, E. F., Zakharchenya, B. P., and Reinov, N. M., 1953, *Dokl. Akad. Nauk S.S.S.R.*, **92,** 265 ; 1956, *J. tech. Phys., Moscow*, **26,** 700.

Haken, H., 1956 a, *J. Phys. Radium*, **17,** 826 ; 1956 b, *Nuovo Cim.* **3,** 1230 ; 1956 c, *Ibid.*, **4,** 1608 ; 1956 d, *Z. Phys.*, 146, 527 ; 1956 e, *Z. Naturf.*, **11,** 875 ; 1957, *Halbleiterprobleme*, Edited by W. Schottky (Braun-, schweig : Vieweg), **4,** in the press.

Hayashi, M., and Katsuki, K., 1950, *J. phys. Soc., Japan*, **5,** 380 ; 1952, *Ibid.*, **7,** 599.

Heller, W. R., and Marcus, A., 1951, *Phys. Rev.*, **84,** 809.

Hilsch, R., and Pohl, R. W., 1928, *Z. Phys.*, **48,** 384.

Kokhanenko, P. N., 1954, *J. exp. theor. Phys.*, **26,** 121.

Kokhanenko, P. N., and Gol'tsev, V. D., 1952, *Dokl. Akad. Nauk S.S.S.R.*, **85,** 543.

Landolt-Börnstein, 1955 (6th Edition ; *Berlin, Göttingen, Heidelberg :* Springer-Verlag), Vol. 1, pt. 4, p. 869.

Martienssen, W., 1957, *J. Phys. Chem. Solids*, **2,** 257.

Meyer, H. J. G., 1956, *Physica*, **22,** 109.

Mott, N. F., and Gurney, R. W., 1948, *Electronic Processes in Ionic Crystals* (Oxford : Clarendon Press).

Muto, T., and Okuno, H., 1956, *J. phys. Soc., Japan*, **11,** 633.

Nikitine, S., 1955 a, *J. Phys. Radium*, **16,** 40 ; 1955 b, *Helv. phys. acta*, **28,** 308 ; 1956, *J. Phys. Radium*, **17,** 817.

Nikitine, S., Couture, L., Perny, G., and Reiss, R., 1955, *C.R. Acad. Sci., Paris*, **241,** 629.

Nikitine, S , Couture, L., Sieskind, M., and Perny, G., 1954, *C.R. Acad. Sci., Paris*, **238,** 1786 ; 1955, *J. Phys. Radium*, **16, 41**, 42.

Nikitine, S., Couture, L., Perny, G., Sieskind, M., and Reiss, R., 1955, *Supplément au Bull. Inst. Intern. du Froid*, Annexe 3.

Nikitine, S., and Perny, G., 1955 a, *C.R. Acad. Sci., Paris*, **240, 64** ; 1955 b, *Ibid.*, **240,** 2298 ; 1955 c, *J. Phys. Radium*, **16, 136** ; **1956,** *Ibid.*, **17, 1017** ; **1958 (in the press).**

Nikitine, S., Perny, G., and Sieskind, M., 1954 a, *C.R. Acad. Sci., Paris*, **238,** 67 ; 1954 b, *Ibid.*, **239,** 247 ; 1954 c, *J. Phys. Radium*, **15,** 18.

Nikitine, S., and Reiss, R., 1956 a, *C.R. Acad. Sci., Paris*, **242,** 238 ; 1956 b, *Ibid.*, **242,** 1003 ; 1956 c, *J. Phys. Radium*, **17,** 1017 ; 1957 a, *Ibid.*, **18,** 74 ; 1957 b, *C.R. Acad. Sci., Paris*, **244,** 1478 ; 1957 c, *Ibid.*, **244,** 2788 ; 1957 d, *Ibid.*, **245,** 52.

Nikitine, S., Reiss, R., and Perny, G., 1955, *C.R. Acad. Sci., Paris*, **240,** 505 ; 1956 a, *Ibid.*, **242,** 1588 ; 1956 b, *Ibid.*, **242,** 2540.

Nikitine, S., and Sieskind, M., 1955, *C.R. Acad. Sci., Paris*, **240,** 1324.

Nikitine, S., Sieskind, M., and Perny, G., 1954, *C.R. Acad. Sci., Paris*, **238,** 1987.

Overhauser, A. W., 1956, *Phys. Rev.*, **101,** 1702.

Peierls, R., 1932, *Ann. Phys. Lpz.*, **13,** 905.

Perny, G., 1958, *Thesis*, University of Strasbourg (in the press).

Perny, G., and Nikitine, S., 1957, *C.R. Acad. Sci., Paris*, **244,** 878.

Reiss, R., 1956, *C.R. Acad. Sci., Paris*, **243,** 902.

Schneider, E. G., and O'Bryan, H. M., 1937, *Phys. Rev.*, **51,** 293.

Sieskind, M., 1956, *J. Phys. Radium*, **17,** 821 ; 1957, *C.R. Acad. Sci., Paris*, **245,** 1006.

Sieskind, M., and Nikitine, S., 1957, *C.R. Acad. Sci., Paris*, **245,** 659.

Wannier, G. H., 1937, *Phys. Rev.*, **52,** 191.

Zollweg, R. J., 1955, *Phys. Rev.*, **97,** 288.

Crystallinity Effects in the Electron Microscopy of Polyethylene†

By A. W. Agar
Aeon Laboratories, Egham, Surrey

F. C. Frank and A. Keller
H. H. Wills Physics Laboratory, University of Bristol

[Received July 16, 1958]

Abstract

Single crystals of polyethylene were examined electron-microscopically with special precautions to avoid damage due to the electron beam. A variety of unexpected beam-sensitive features were observed. They can be classified as (1) moirés, (2) morphological features. In connection with (1) the general principles of moirés are discussed. It is shown that moirés can be due either to double diffraction, the conditions of the formation of which are defined, or to dislocations arising through contact deformation of superimposed crystals. Our observations on polyethylene crystals are analysed in the light of this general discussion. It is found that the moirés in polyethylene are to a large extent due to rotational superposition of layers. Crystal growth with such a rotation is exemplified by a unique preparation. In places the moiré makes a pseudo-image of the crystal lattice itself, in others only of particular lattice planes. There is evidence for strong local lattice distortions, and also for possible local departures from orthorhombic symmetry. Various other features are also discussed and fundamental difficulties in the interpretation of image contrast are pointed out. (2) We found a surface corrugation strongly heightened by extinction effects which disappeared within a short time after the preparation of the specimens. We tentatively interpret this as resulting from a change in the molecular fold length, when the crystals are dried down. The corresponding striations did not follow low index crystallographic directions. There is evidence for the existence of distinguishable quadrants within the crystals, and also of internal buckling, both of which are consistent with other considerations. Repeated asymmetric twinning along (010) and to a lesser extent along (100) is observed which can account for a variety of observed crystal shapes.

§ 1. Introduction

It has been found by one of us (Keller 1957) and also by Till (1957) and by Fischer (1957) that the linear polyethylene Marlex 50 can be obtained in the form of single crystals by lowering the temperature of a solution. The crystalline nature of the product was revealed by the often regular lozenge-shaped habit and also by the electron diffraction patterns obtained from individual crystals within the electron microscope. According to the

† Communicated by the Authors.

diffraction patterns the molecules were perpendicular to crystal layers of thickness about 100 Å, and it was concluded that the long molecules must sharply fold back on themselves in order to take up this configuration (Keller 1957). Low angle x-ray results gave further support to this deduction (Keller and O'Connor 1957). The electron diffraction patterns were unusually sharp, but rapidly deteriorated and finally disappeared under the effect of electron bombardment within the microscope (Keller 1958), while the shape of the crystals remained unaltered. This could only be avoided, or rather retarded, by working with very low beam intensities, which made focusing of images practically impossible at any but the lowest powers. All our previous electron micrographs show objects which have lost their crystallinity in this way at the time when the photographs were taken, and it is very probable that this applies to most other electron micrographs of polymers, and possibly also of many other organic substances taken in the past. Although we had no evidence of any morphological change in consequence of this loss of lattice order, we found it unsatisfactory in principle that the crystals could not be viewed in their original ordered state. The present work was undertaken in order to overcome this difficulty with the available experimental facilities.

§ 2. Experimental

2.1. *Material and Sample Preparation*

The material used in these investigations was the linear polyethylene Marlex 50. Thin films were dissolved in boiling trichlorethylene. When the solution was cooled to room temperature the polymer precipitated forming a suspension of lozenge-shaped crystals, which were representative of the whole material. One other preparation will also be referred to. This was a fraction of Marlex obtained by Soxhlet extraction with trichlorethylene boiling at a reduced pressure at 80°C. (The extraction experiments were part of another programme of work.) On cooling the solution containing the extract, unusual crystals were obtained which were found to be relevant to the present work. In all cases the crystals were deposited on carbon supporting films.

2.2. *Method*

The specimens, dried down on to carbon-coated specimen grids were examined in a Siemens Elmiskop I, and in some cases in a Philips instrument. The specimens were known to be very prone to change in the electron beam, and a modified operating procedure was therefore adopted. The double condenser lens system in the Siemens microscope enabled the illumination to be restricted to a spot only about 3 microns in diameter when the condenser was focused. Crystals moved into the beam under these conditions were seen to change a fraction of a second after entering the beam, all internal structure disappearing (the crystal outline remaining unchanged). When the second condenser lens was

defocused, the illumination spread over some 50–100 μ, i.e. about one square of the supporting grid, and at this intensity level, the crystals could be irradiated for perhaps half a minute before starting to deteriorate markedly. At this level of illumination it was not possible to focus the picture. Consequently a crystal was focused carefully at a relatively high intensity, and after the condenser had been defocused, the specimen stage was moved a minimum of two grid squares to bring a hitherto unirradiated crystal into view (the intensity was just sufficient to recognize a crystal outline). An exposure of about 15 seconds was found to yield a plate of reasonable density. It was found that photographs could be recorded at a magnification of 5000 with a high percentage of success, of a quality suitable for further optical enlargement of 5–15 times. Working to these limits, the movement of the specimen stage was not found to affect the focus of the picture, and the stability of the H.T. and lens current supplies was quite adequate. Higher magnifications were not found to yield a sufficient proportion of good pictures to be worth pursuing; the increased exposure to the beam required usually resulted in a deterioration in the crystal even when a well-focused picture was obtained (and this proved very difficult).

2.3. *Experimental Results*

The photographs revealed a variety of new features which normally were all destroyed by the electron beam. As seen in fig. 1, Pl. 6 there are striations, in places forming cross gratings, in most areas where the crystal was more than one layer thick. The striation is along different directions, is of variable spacing and is often curved. Figures 2–7, Pls. 7–9 show various details of such patterns to be discussed later. Figures 8 and 9, Pl. 10 are a pair of dark and light ground photographs taken of the same area, the light ground having been taken first.

In some photographs there was an additional feature in the form of a rather irregular line-system following roughly but not exactly the outlines of the crystals (fig. 10, Pl. 11 and also fig. 12, Pl. 12). It appears as if the crystal were divided into four quadrants. At first sight the cause of this structure is seen to be an extinction effect but closer examination reveals also the presence of protrusions on the surface as suggested by the enlarged central portion of the above photograph (fig. 11, Pl. 12 and also by fig. 12). The appearance of this type of structure was erratic. We finally established that such structures are only to be found in crystals examined immediately after their preparation for electron microscopy. Once dried on the supporting film the specimen loses this structure within a very few hours even without being placed into the microscope. After the obvious structural features have disappeared extinction bands could still be seen running parallel to the original structure (fig. 13, Pl. 13). These bands broadened and finally disappeared within a few hours (e.g. fig. 9). Although the structures described in this paragraph were even more difficult to record than those in figs. 1–7 in some cases they persisted

even when the latter and all other obvious extinction effects had disappeared (fig. 14, Pl. 13). In one instance we succeeded in shadowing these rapidly changing specimens while still possessing a structure as in fig. 14. The shadowing (which happened to be too light to merit a separate illustration here) confirmed the existence of protrusions.

The new features shown by figs. 1–14 are affected by the electron beam. However, the two types of effects (fig. 1–7 and 10–14) are not closely interconnected as each may exist separately after the other has already disappeared.

The crystals formed from the extract solution mentioned above were also lozenge-shaped, but in addition they revealed a quadruple leaf pattern, a double leaf along each diagonal (fig. 15, Pl. 14). Sometimes half of the pattern was missing leaving an S shaped line along each diagonal. It can be seen that the line producing the leaf pattern is in fact the envelope of the apices of successive layers, the leaf shape arising through the successive layers not being parallel but being rotated with respect to each other, always in the same sense through approximately the same angle. It can be seen from the shadowing that one half of the figure is formed on the top of the crystal, i.e. above the largest first layer while the other half results from the succession of layers in the opposite direction. The figure S referred to above appears in crystals which only thickened in one direction. It is noticeable in fig. 15 that lozenges in successive layers change their shapes continuously; the long diagonal of the outer layers corresponds (by continuity along the S-curves) to the short diagonal of inner layers. These crystals were generally too thick for electron diffraction or even for detailed direct viewing. Nevertheless diffraction patterns could be obtained of the thinner corners. Figure 16, Pl. 14 shows that each strong spot has a tail consisting of regularly spaced spots. When photographed in the Elmiskop with the aid of the double focus condenser as figs. 1–14 striations could be found at the thin edges (fig. 17, Pl. 15).

In the course of the electron diffraction work on Marlex spots frequently appeared which did not correspond to spacings predicted by the structure and sometimes also curious spikes could be seen. As such effects are relevant to the present problem two are illustrated by figs. 18 and 19 (*a*) and (*b*), Pls. 15 and 16.

§ 3. Discussion

The present experiments revealed two new features in the electron micrographs which have been unnoticed before owing to the damaging effect of the electron beam. The first one (figs. 1–7) is a direct consequence of the lattice order within the crystal. These represent moiré patterns. The second effect corresponds, at least partly, to a genuine morphology.

3.1. *Moiré Patterns*

3.1.1. *The formation of moiré patterns*

Moiré patterns may be, and frequently are, produced in electron microscopy whenever thin crystals with small differences in orientation or

lattice spacing are superposed (Pashley *et al.* 1957). Two modes of formation of these patterns quite distinct in principle are involved. These are:

(a) double Bragg diffraction, successively on the two crystals, and

(b) mutual deformation at the surface of contact, most succinctly described as the formation of a grid of dislocations.

They are clearly distinct because mode (a) is able to generate moiré patterns in the absence of actual contact between the crystals, which is a necessary condition for (b); while (b) can produce a visible pattern of similar character when the conditions for single Bragg diffraction (an evident pre-requisite for double Bragg diffraction) would not be satisfied in the absence of the mutual deformations. The fact that dislocations singly or in grids make a sufficient localized lattice strain to be directly visible by electron microscopy, has been demonstrated for metallic films by Hirsch *et al.* (1956). However the kind of contrast produced in the image of dislocations changes as a Bragg extinction setting is approached. For this reason, though in principle the two modes of moiré formation are distinct because either may exist in the absence of the other, nevertheless the likely real situation when we deal with crystals in true contact is one in which the mode of formation of the moiré pattern is not sharply assignable to (a) or (b); and it is probably only as an approximation that it may be regarded as a mere additive combination of the two. Nevertheless, we shall so regard it in this discussion.

The patterns formed contain similar periodicities by either mode of formation: namely 'vernier periods' between the two crystal lattices.

Moiré patterns are formed on a more macroscopic scale by the superposition of pairs of patterned masks—e.g. two similar or near similar pieces of woven silk—when the mode of formation may be interpreted without recourse to diffraction theory. It is necessary to realize that this is not a distinct mode of formation of the pattern, but an alternative description of the double diffraction mode permissible in the limit when we have two dimensional patterns on a scale large compared with the wavelength. According to this superposition theory, the transmission function $F_{12}(x, y)$, for two superposed masks is the product $F_1 . F_2$ of the two separate transmission functions. However, the moiré pattern in F_{12} is not properly seen by too close an inspection, which still reveals the structure details of the two layers of woven silk. The moiré pattern is isolated by observing with a lower resolution, i.e. mathematically speaking, by forming the Fourier transform, G_{12} from F_{12}, discarding or diminishing the higher terms from it to give the truncated transform G'_{12}, and re-transforming to F'_{12} which is a lower resolution representation of F_{12}. Now, since F_{12} is equal to the product $F_1 . F_2$, G_{12} is equal to the fold $G_1 \star G_2$ of the separate Fourier transforms of F_1 and F_2. The operation of diffraction of a simple primary beam at the first mask generates a complex of beams corresponding to G_1. The second diffraction of this complex of beams at the second mask generates a complex of beams corresponding to

the fold $G_1 \star G_2 = G_{12}$. Truncation to G''_{12} is performed by the finite aperture of the objective lens (or by the rapidly increasing spherical aberration with large apertures) which forms the image corresponding to F''_{12}.

Thus in the special case considered, as one could expect, the analysis by 'superposition theory' which is geometrical optics, agrees in result with the analysis by 'double diffraction theory', which is wave optics. In other cases, where they differ, wave optics must be preferred.

The foregoing example dealt with a pair of two-dimensional masks ('cross gratings'). In this case diffraction at one setting, from a single monochromatic primary beam, yields a complex of diffracted beams representing the whole Fourier transform. This is not the case for a three dimensional crystal, having a definite thickness: then the same diffraction experiment yields only a part of the Fourier transform—that which lies on the surface of the Ewald sphere in reciprocal space. Now the superposition theory is not applicable in general. However, it remains applicable with some modifications, for important classes of specimens, including our own. With high energy electrons, crystals of moderate thickness yield cross-grating patterns whenever the incident beam is parallel, or nearly parallel to a prominent zone axis. Conditions for moiré formation by double diffraction are then likely to be satisfied if the two crystals have this zone axis in common. 'Superposition theory' may then be used for a quick simple interpretation of the moiré pattern formed, but not by the superposition of real layers of atoms, but of fictitious layers each of which is the Fourier transform of the reciprocal lattice plane to which the Ewald sphere is tangential. In the general case this does not correspond to any actual plane of atoms in the crystal. However the conditions we speak of are most readily produced when the crystal symmetry and habit provide a tabular development on a plane normal to a zone axis. Then it occurs naturally that superposed thin crystals have this zone axis in common, and parallel to a probable direction for the incident beam. It also follows that there are real atomic layers in the crystal parallel to the diffracting plane in reciprocal space, and that these layers have a pattern which, though not necessarily the same as that of the Fourier transform of the reciprocal lattice plane is rationally related to it. Then the double diffraction moiré and the dislocation pattern at the interface will be closely related to each other.

Let vectors $\mathbf{k}_{1,i}$ be reciprocal lattice vectors of the first crystal, i.e. Fourier components of the potential in this crystal, so that $|\mathbf{k}_{1,i}| = 1/d_{1,i}$ where $d_{1,i}$ is an interplanar spacing in this crystal, and the direction of $\mathbf{k}_{1,i}$ is normal to the corresponding planes: and let $\mathbf{k}_{2,j}$ be reciprocal lattice vectors of the second crystal. Then the Fourier components of the moiré pattern are necessarily of the form

$$\mathbf{k}_{3l} = \mathbf{k}_{1,i} - \mathbf{k}_{2,j}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Which actual vectors $\mathbf{k}_{1,i}$ and $\mathbf{k}_{2,j}$, if any, take part in the formation of the moiré image depends on the illumination conditions in a way which we

shall examine presently; and in any case only the *small* difference vectors $\mathbf{k}_{3i}$ are accepted. A simple case is that in which the two crystal lattices are parallel, but differ slightly in the spacings, $d_{1,i}$ and $d_{2,i}$ of corresponding diffracting planes. Then

$$|\mathbf{k}_{3,i}| = |\mathbf{k}_{1,i} - \mathbf{k}_{2,i}| = 1/d_{1,i} - 1/d_{2,i}.$$

The resulting periodicity in the moiré pattern is parallel to that in the crystals which causes it, but of longer spacing,

$$d_{3,i} = 1/|\mathbf{k}_{3,i}| = d_{1,i}d_{2,i}/(d_{1,i} - d_{2,i}). \quad . \quad . \quad . \quad . \quad . \quad (2)$$

A second simple case is that in which the spacings are the same, but the two crystals have a small relative rotation α. Then the contribution to the moiré pattern from two corresponding periodicities $\mathbf{k}_{1,i}$, $\mathbf{k}_{2,i}$ (for which $|\mathbf{k}_{1,i}| = |\mathbf{k}_{2,i}| = 1/d_i$) has the spacing

$$d_{3,i} = 1/|\mathbf{k}_{3,i}| = d_i/(2 \sin \tfrac{1}{2}\alpha) \sim d_i/\alpha \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and is orthogonal to the mean direction of the corresponding periodicities in the two crystals.

In the simplest basic cases moiré fringes of the first case correspond with the edge dislocations at the interface between the crystals of different lattice spacing, those of the second case with the screw dislocations at the 'twist boundary' interface. If we were dealing only with monolayers, the correspondence would be complete. With thicker crystals, the nature of the correspondence depends on the particular selection of diffractions which actually contribute to moiré formation: there may for example be an integral number, different from unity, of moiré fringes per dislocation.

3.1.2. *Conditions for the formation of double diffraction moirés*

3.1.2.1. *General considerations*

The genesis of double diffraction moirés can be easily demonstrated by considering diffraction at the first crystal and by following the diffracted beams through the second crystal. This discussion is conducted in terms of the 'kinematic' theory of diffraction. We estimate that in these hydrocarbon crystals, even for the strongest reflections, the 'extinction distance' is of the order of 1000 Å. The thickness of the crystal layers we are concerned with is sufficiently smaller than this to justify the employment of 'kinematic' rather than 'dynamic' theory.

A beam diffracted once can be the source of further diffracted beams when entering the second crystal. If the second crystal is exactly identical with the first and is in the same orientation, then the original diffraction pattern (or at least that part of it which still satisfies the Laue conditions) has to be repeated around the corresponding first diffraction spot. Obviously one of the secondary diffracted beams will coincide with the original primary beam. If, however, the second crystal is slightly different in either a lattice spacing or in its orientation the same secondary beam which in the previous case coincided with the original primary beam

will now be slightly displaced and produce a separate reflection near the origin. By taking other primary reflections the whole diffraction pattern can be repeated around the origin on a reduced scale. Figure 18 illustrates this point. We see six spots around the origin with the same symmetry as the primary 110 and 200 spots. If only these diffraction spots are used in the image formation, a moiré pattern will result which is essentially an enlarged image of the original lattice. This selection of the innermost secondary diffraction spots (truncation of the Fourier transform G_{12} in terms of the nomenclature in the preceding section) is achieved by the objective aperture of the microscope or by the spherical aberration of the objective. The secondary diffraction spots in fig. 18 arise through four superposed twins, having the twin planes of the type {110} and a composition plane (001). They correspond to spacings ~ 10 Å, hence the resulting moirés would still not be resolvable by our photographs. However, it is easy to see that in the same way spots could arise which are still nearer to the origin thus corresponding to still larger spacings.

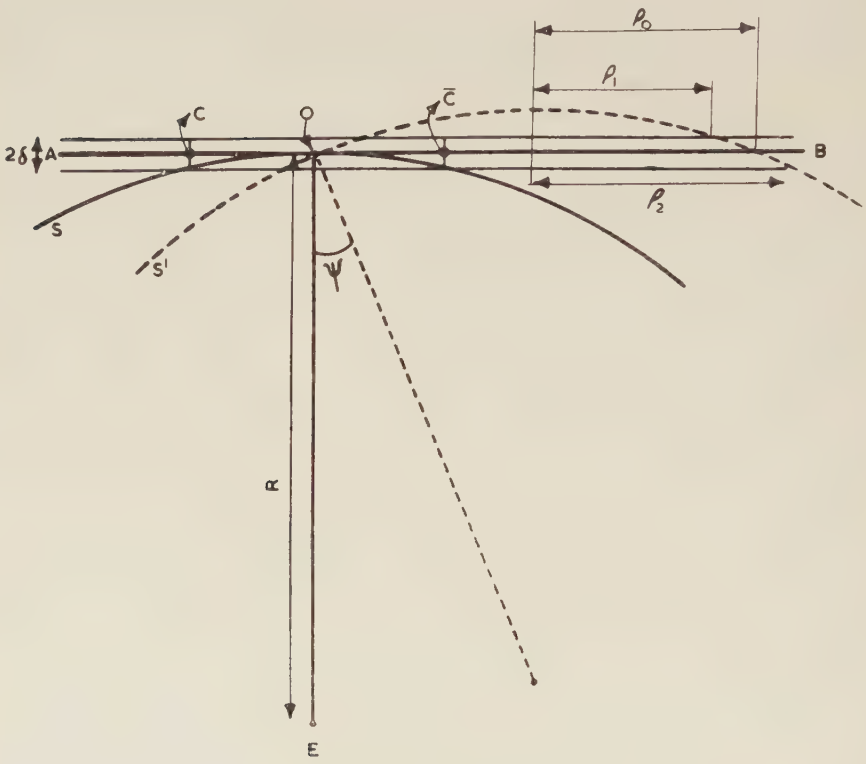
Additional spots will also appear within the immediate neighbourhood of the primary reflections due to the first crystal. These can be due to secondary diffraction just as around the primary beam but they could also arise through primary diffraction from the slightly different second crystal. If a region around such a reflection is selected by displacing the aperture (as done in dark field work) the same type of moirés should appear as when the close environment of the incident beam only is selected.

If the primary beam and one of the strong reflections are both admitted a dark and light field image will be superimposed. However, the lens aberrations in the angular region of a usual Bragg reflection are appreciable and the image formed by rays passing through such an outer part of the lens will be displaced with respect to that formed by the central rays. The result will be a usual bright field and a weaker displaced dark field image. The direction of the displacement is normal to the planes producing the reflection responsible for the dark ground image. The moirés will be similarly displaced, and thus in parts they can be superimposed with either reinforcement or reduction of contrast.

As double diffraction is the pre-requisite of this type of moiré formation conditions of repeated diffraction will be examined more closely. In the first place the requirements which have to be satisfied for a particular primary diffraction to occur will be defined. Such diffractions will occur regardless whether the diffracted beams are accepted by the aperture or modified otherwise by the imperfections of the following optical system of the microscope. The beam can be considered as perfectly collimated owing to our low condenser setting. We take a crystal which has one principal zone axis approximately parallel to the beam (a condition which is satisfied for our polythene crystals). At first we take this zone axis as exactly parallel to the beam. In fig. 20 we view the reciprocal lattice plane $\overline{\mathrm{AB}}$ edge on. O is the origin of the reciprocal lattice, C and $\overline{\mathrm{C}}$ are a pair of lattice points. The beam EO is perpendicular to the $\overline{\mathrm{AB}}$ plane,

i.e. parallel to the corresponding zone axis. S is the Ewald sphere. Diffraction will occur whenever a reciprocal lattice point intersects the Ewald sphere. It is seen that this condition cannot be satisfied for any lattice point in the $\overline{AB}$ plane. If the crystal were two-dimensional, each reciprocal lattice point would be extended into an infinite line perpendicular to the $\overline{AB}$ plane. These lines would intersect the Ewald sphere

Fig. 20



Construction to illustrate conditions of reflection from a reciprocal lattice layer of thickness 2γ with the beam normal to the layer and with the beam at an angle ψ off the normal position.

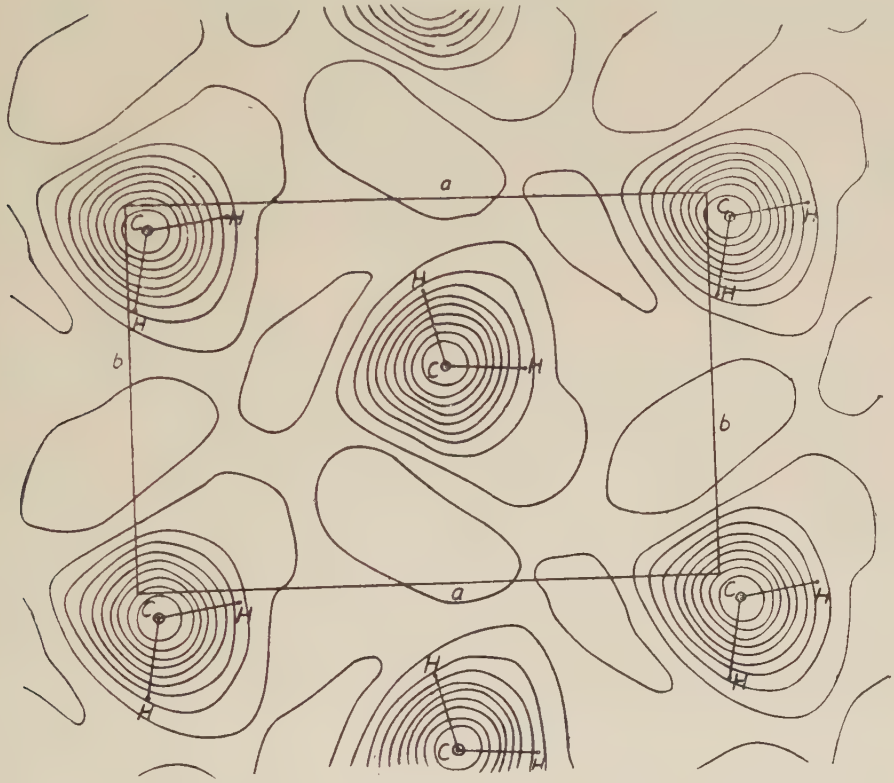
and thus conditions for diffraction would be satisfied. If the crystal is not truly two-dimensional but still very thin, lattice points C, $\bar{C}$ etc. will have to be replaced by lines of finite length. Thus, whether diffraction is possible or not will depend on the length of the line, i.e. on the thinness of the crystal. Conditions for diffraction from the $\overline{AB}$ reciprocal lattice plane may also be satisfied if the crystal is tilted, i.e. when the $\overline{AB}$ plane

is not exactly perpendicular to the beam. Thus diffraction from the $\overline{AB}$ reciprocal lattice plane will occur either if the crystals are sufficiently thin or if they are suitably inclined to the beam.

The spread of the reciprocal lattice point due to the thinness may be assessed from the formula

$$I = \sin^2(\pi N\zeta)/\sin^2(\pi\zeta) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Fig. 21



Electron density distribution in a polyethylene unit cell in c projection (after Bunn).

where I is the interference function which defines the intensity distribution along the reciprocal lattice direction c^*, ζ is the displacement from a Bragg point expressed in units of the reciprocal lattice vector in this direction, and N is the number of repetitions along the crystal axis c. N being larger than 2, this falls to a little less than half its maximum value when $\zeta = 1/2N$, and then rapidly to smaller values, remaining small until the neighbourhood of the next reciprocal lattice point. We accordingly take this as the effective extension of the reciprocal lattice point, and

replace the plane AB in fig. 21 by a layer of thickness $2\delta = 1/N$, in terms of the reciprocal lattice spacing c^*.

Further we tilt the crystal through an angle ψ which can be represented by tilting the Ewald sphere through this angle around O. S′ the tilted sphere and the corresponding radius to O are drawn with dotted lines. It is seen that in general the tilted sphere and the reciprocal layer will intersect along a ring, having inner and outer radii ρ_1 and ρ_2 respectively, given by

$$\rho_1^2 = R^2 - (R\cos\psi + \delta)^2 \text{ and } \rho_2^2 = R^2 - (R\cos\psi - \delta)^2.$$

If δ is small as in our case (see below) δ^2 can be neglected and we have

$$\rho_1^2 = (R\sin\psi)^2 - 2R\delta\cos\psi,$$
$$\rho_2^2 = (R\sin\psi)^2 + 2R\delta\cos\psi. \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

If ψ is small ρ_1 is imaginary, in which case sphere and layer will intersect in a full circle. The tilt when this full circle becomes hollow in the centre is reached when $\rho_1 = 0$ which defines the angle

$$\psi = \cos^{-1}(1 - \delta/R). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

As ψ increases the first terms on the right-hand side of (5) become large in comparison with the second terms and ρ_1, ρ_2 will approach the values $R\sin\psi(1 - \delta/R\sin^2\psi)$ and $R\sin\psi(1 + \delta/R\sin^2\psi)$ respectively, in the range where $\cos\psi$ is still near 1. In this case $\rho_2 - \rho_1 = \Delta\rho = 2\delta/\sin\psi$.

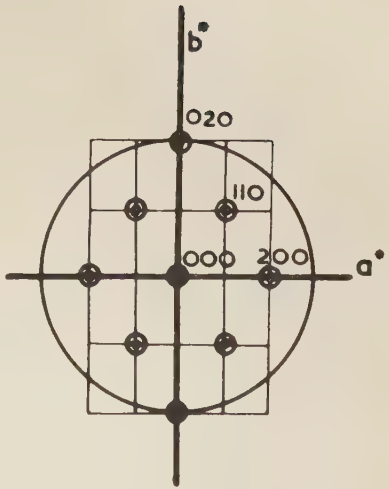
The centres of the circles are displaced from O through a distance $\rho_0 = R\sin\psi$.

In this way knowing δ and ψ the circle or ring of intersection of sphere and reciprocal lattice can be defined and by finding the reciprocal lattice points within this area of intersection conditions for primary diffraction can be determined. If the diffracted beams are diffracted again by a second crystal the same procedure has to be repeated for each diffracted beam. Each diffracted beam has to be taken as a primary beam, and in this way conditions for secondary diffraction can be laid down. However the procedure can be largely simplified as, in the case of our moirés, conditions for secondary diffraction are automatically satisfied whenever a primary diffraction can take place as will be shown in § 3.1.2.2 where the above deductions will be applied to polyethylene.

3.1.2.2. *Conditions for double diffraction moirés in polyethylene crystals*

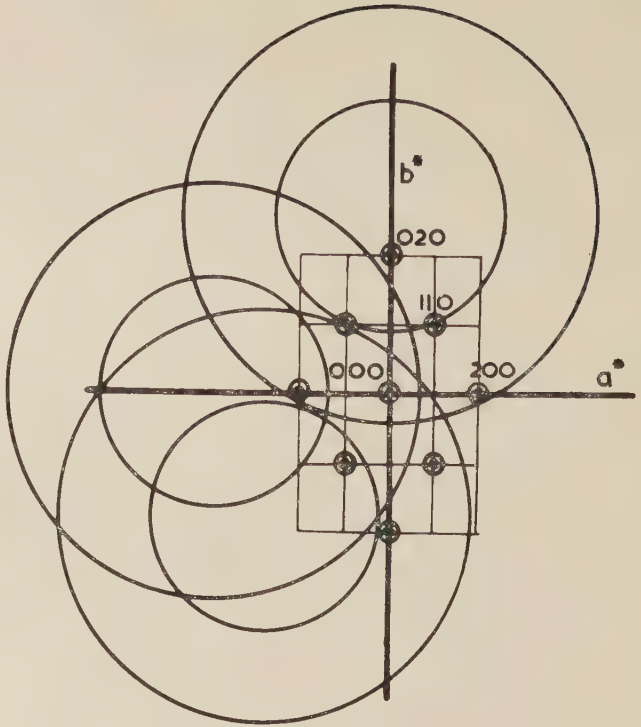
The lattice of polyethylene is orthorhombic with $a = 7{\cdot}40$, $b = 4{\cdot}96$ and $c = 2{\cdot}53$ Å with two molecules per unit cell (Bunn 1939). In our photographs we view the crystals along the c-axis, which is also the direction of the molecules. Figure 21 shows the electron density distribution in this projection as determined by Bunn (1939). All strong reflections are of the form $hk0$ and consequently it is sufficient to consider only the $hk0$

Fig. 22



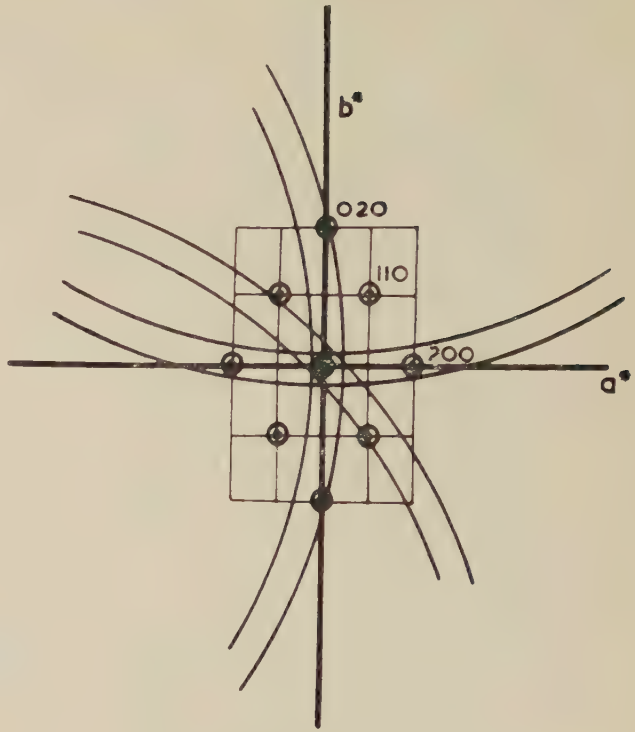
Beam parallel to c.

Fig. 23



Beam at 1° 30′ to c.

Fig. 24



Beam at 5° to c.

Regions of reflections within the $hk0$ reciprocal lattice layer in c projection in a polyethylene crystal 120 Å thick, with a 80 kv electron beam.

reciprocal lattice plane as this only will contribute appreciably to the diffraction pattern, hence to the moirés. The beam is approximately perpendicular to this reciprocal lattice plane, owing to the tabular crystal habit and orthorhombic symmetry.

The layers of which the crystals are built are about 120 Å thick. This corresponds to about 50 repetitions along the c-axis. Accordingly the effective thickness of a reciprocal lattice plane 2δ may be taken as $1/(50c) = 1/125$ Å. The table gives ρ_0, ρ_1, ρ_2 and $\Delta\rho = \rho_2 - \rho_1$ for increasing angles of tilt. For a 80 kv electron beam $R = 20\,\text{Å}^{-1}$.

ψ	0	$\frac{1}{2}°$	1°	1° 30′	2°	3°	5°	10°
ρ_0	0	0·174	0·46	0·52	0·70	1·04	1·74	3·47 Å^{-1}
ρ_1	0	—	—	0·33	0·57	0·96	1·70	3·48 Å^{-1}
ρ_2	0·40	0·43	0·530	0·61	0·82	1·12	1·79	3·44 Å^{-1}
$\Delta\rho$	—	—	—	0·28	0·25	0·16	0·09	0·04 Å^{-1}

Figure 22 shows the $hk0$ reciprocal lattice plane with 110, 200 and 020 reciprocal lattice points which correspond to the most important reflections. The area within the circle represents the region within which the sphere of reflection lies within the reciprocal layer of thickness 2δ. All significant reflections are within this area except the weakest, 020, which is on the periphery. With increasing tilt the circle shifts and becomes a ring when $\psi = 1° 9'$ and at this stage some of the reciprocal lattice points shown by fig. 22 will not lie within the region of intersection any longer. Figure 23 shows the rings of intersection for $\psi = 1° 30'$ for three different tilt-directions. As seen, the region of intersection only includes two to four reciprocal lattice points, in addition to the origin. With increasing tilt the rings become narrower, and the number of low order points included fewer. If ψ is greater than 3°, there may be none, apart from the origin (i.e. no diffraction) or there may be one $hk0$, probably with its symmetrical partner—($hk0$) as well. Figure 24 shows the same construction for $\psi = 5°$.

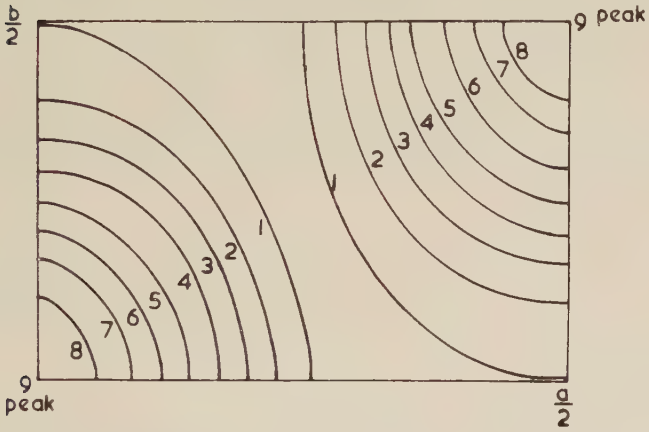
Above we have discussed the conditions for primary diffraction. Now we may ask what happens when the diffracted beam passes through a second crystal which is slightly different in orientation, in lattice spacing or in both. The moiré fringes we see are spaced at distances of 50 Å or more. The corresponding reciprocal lattice points are at a distance of $0{\cdot}02\,\text{Å}^{-1}$ or less from the origin. As seen from the table and from figs. 22–24, the ring of intersection is wider than this even for tilts of 10°, consequently it will practically always include the reciprocal lattice point corresponding to the moiré we can observe. Thus whenever a primary

diffraction can take place conditions for double diffraction –giving rise to moirés on the scale observed—are also satisfied.

By Abbe's theorem the images formed in the microscope result from the recombination of the diffracted beams (including also the primary beam). The moirés are magnified imperfect images of the lattice formed by these doubly diffracted beams which are nearly parallel to the primary beam. Next we want to know what type of image to expect when only a limited number of diffracted beams take part in the image formation. This can be determined by forming the squared Fourier transform of the array of the reciprocal lattice points concerned, weighted according to the intensities of the corresponding reflections. When taking only one diffracted beam its recombination with the primary beam of equal intensity will lead to the image of a line grating with an intensity variation given by

$$I = 4\cos^2 2\pi x. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Fig. 25



Density contours in an image formed through the recombination of the 110, $\bar{1}10$ and 000 reflections, all being of equal intensity. One-quarter of the unit cell.

(For the sake of convenience the coordinates of the two reciprocal lattice points were taken as 1, 0, 0 and -1, 0, 0.) If the two intensities are dissimilar, in the ratio $n:1$ the corresponding expression will be

$$I = (n-1)^2 + 4n\cos^2 2\pi x. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Thus the difference in intensities will not affect the periodic nature of the image nor the periodicity itself, but it diminishes the contrast through a uniform background. Above we utilized only one reflection (say *hkl*). Inclusion of the symmetrical pair $-(hkl)$ will improve the contrast by increasing the amplitude variation but otherwise will not alter the image. For this reason symmetrical pairs are counted as one reflection for the present.

We proceed by including a second reflection. Let us take the strongest 110 and $1\bar{1}0$. For the sake of convenience we chose the origin half way along the line connecting the 110 and $1\bar{1}0$ reciprocal lattice points. Then the three lattice points 110, $1\bar{1}0$ and 000 will have coordinates 1, 0, 0, -1, 0, 0, and 0, 1·5, 0, respectively. The corresponding Fourier transform is given by

$$T_{xy} = \exp(2\pi i x) + \exp(-2\pi i x) + \exp[2\pi i(1{\cdot}5y)]$$
$$= 2\cos 2\pi x + \cos 2\pi(1{\cdot}5y) + i \sin 2\pi 1{\cdot}5. \quad . \quad . \quad . \quad . \quad (9)$$

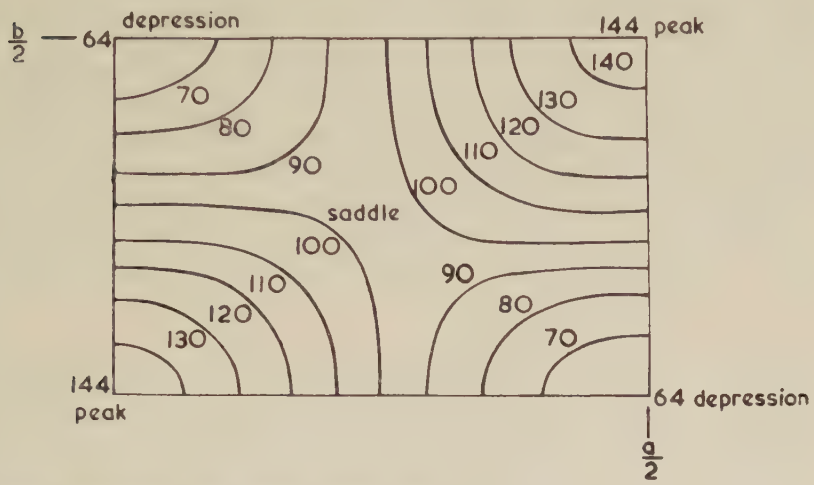
After multiplication with its complex conjugate the transform is plotted in fig. 25. The plot only extends over one quarter of the unit cell of polyethylene. Comparison with fig. 21 reveals that the essential features of the lattice are clearly represented as we have well defined peaks where there are atoms. For the thicknesses involved, the primary beam may be stronger than even the strongest diffracted beam. In this case the corresponding terms in the Fourier transform will have to be weighted accordingly. Tentatively we took the primary beam 10 times stronger than the 110 reflections. The corresponding transform is shown by fig. 26. Again the same peaks are visible as before except that they are superimposed on a considerable background which would lead to lower contrast in the image. Further there is a noticeable saddle along the [110] direction. This will produce streaks along [110] between the peaks in the image.

Summing up: recombination of one reflection with the primary beam gives the image of a line grating. Inclusion of a second strong reflection is sufficient to reproduce the image of the lattice in the particular projection, possibly with more or less pronounced streaks connecting the individual peaks.

At this stage the following question arises: Do the peaks in figs. 25 and 26 represent electron surplus or electron deficiency, i.e. will the atoms be 'dark' or 'light' in the photographs? This requires a fundamental reconsideration of the origin of contrast in electron microscope images. Even in the case of *direct* imaging of atoms, the final contrast will depend on a number of factors. In the first place the phase problem in electron scattering by atoms is not a simple one. It involves consideration both of 'simple' elastic scattering and the elastic scattering component associated with inelastic scattering, attributable to the 'gap' in the primary wave-front which this produces (see, e.g., Haine 1957), and also of the second-order phase corrections for 'simple' elastic scattering, dependent on nuclear mass (Glauber and Schomaker 1953). To our belief there is no complete theory of this scattering process which includes all relevant factors sufficiently for a precise analysis of image-contrast. Thus the contrast of atoms, i.e. whether they will be light or dark in a *direct* image, cannot be predicted with any confidence. In the case of moiré representation of atoms there will be an additional phase factor owing to the

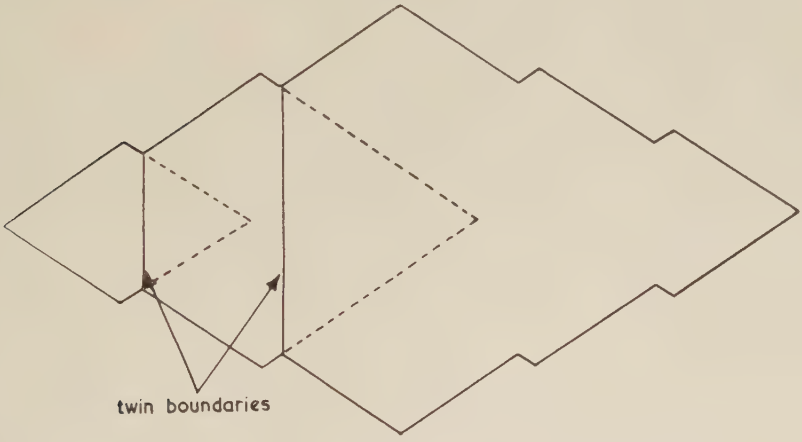
difference between the forward path of the primary beam and the oblique path of a diffracted beam, between the successive diffractions. This phase difference Φ will be added to the effect produced by the elementary

Fig. 26



Density contours in an image formed through the recombination of the 110, $\bar{1}10$ and 000 reflections, the 000 having ten times the intensity of the other two. One-quarter of the unit cell.

Fig. 27



Diagrammatic illustration of repeated twinning.

act of scattering, the latter to be taken twice, because we have repeated diffraction. For the nth order diffraction at spacing d with the thin crystals approximately normal to the beam

$$\Phi = (2\pi D/\lambda)(\sec\theta - 1) \sim \pi D n^2 \lambda / d^2 \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where D is the distance between the mid planes of the two parallel thin crystal plates (the thickness of one of them, if they are equal and in contact) and θ is the angle between direct and diffracted beam. The contrast of direct image not being known, we cannot predict the contrast of the moiré image either. However, it follows from (10) that the contrast of a moiré due to a particular spacing d will depend on λ and D. This means that by altering the voltage and the thickness of the crystals the contrast could be made to vanish or even reverse. This particular point should be amenable to experimental verification. The question of contrast is meaningful only if the moiré lattice is well resolved. When we have only sinusoidal fringes the dark and light bands will be equivalent in any case. Poorly resolved cross gratings would give a sort of chess board effect where a distinction between light and dark areas is still meaningless.

3.1.3. *Comparison with experiment*

As both deformation and double diffraction moirés are expected to be similar, we cannot assign the patterns in our photographs readily to one or the other type. Nevertheless there are some guiding principles.

We know from our diffraction experiments that diffraction and also double diffraction does occur, and further that the appearance of the moirés is in general closely connected with the persistence of the diffraction patterns. For this reason and also because double diffraction is the most obvious source of moiré formation in electron micrographs in any case (Pashley *et al.* 1957) it is certain that such moirés contribute significantly to the patterns we see.

There is a fundamental difference in the way deformation and double diffraction moirés are made visible. Double diffraction moirés are formed through the recombination of diffracted and primary beams, or through the combination of diffracted beams with each other. On the other hand the dislocations in the deformation moirés cause elastic distortions within the lattice around them. Thus in the vicinity of dislocations the lattice may be brought nearer to reflecting position or taken further away from it. The diffracted electrons being lost to the image, the lattice around a dislocation line may appear in sufficient contrast to be rendered distinguishable.

The different sources of contrast in the two types of moiré could in principle provide a method of distinguishing between them. We have carried out a series of preliminary experiments where the size and setting of the objective aperture was altered systematically. The different ways in which the contrast of the various features responds to these changes should help to distinguish between the different imaging mechanisms forming the moirés. Such experiments are rendered very difficult by the rapid deterioration of the specimens during a given series, and mainly for this reason the inherent possibilities of this kind of experimentation have not yet been exploited.

There are indications that double diffraction moirés, although predominant, may not be able to account for all effects observed. We fairly often observe rather widely spaced fringes parallel to [010]. Figure 6 shows such an example with a spacing of about 1000 A. If this is due to rotation, and double diffraction, the 020 reflection is involved, but this is weak, with an intensity of only 0·025 of the 110 reflection, and therefore unlikely to produce such a strong contrast. This may therefore be a case in which fringes arise from a difference in a-spacing between the layers. The required difference is 0·3%, if 200 double diffraction produces the moiré. The fringes appear where no Bragg extinction is visible. It may therefore be that we see the dislocations in this case. They would presumably be half dislocations: then the required difference in lattice spacing is the same. But if we see the dislocations rather than a double diffraction phenomenon, it is also possible that the layers are after all rotated on each other (about 0·2°), but only one set of a crossed grid of dislocations are visible.

We observe moiré patterns with spacings down to 40 Å, indicating rotations of up to 6°, or lattice-spacing differences up to 10%. The latter would produce obvious radial splitting or streaking of spots in the diffraction pattern, which is certainly absent. On the other hand extensions of spots as arcs or often chains of discrete spots, along the Debye–Scherrer circles, corresponding to rotations of up to and beyond 6° (represented by the angular spacing between adjacent spots) is frequently observed. Thus while some of the most widely spaced fringes may possibly be due to variability of lattice spacing, relative rotation of the layers is responsible at least for all those of narrower spacing, and is seen to be a very common phenomenon.

Patterns with asymmetric tails have been found by Wilman (1951) in a number of substances and were interpreted by him as arising from a rotational slip. In our case this effect is due to a systematic growth feature (as it may well have been in his). The decreasing intensity of the tails could be due to the decreasing amount of material in the corresponding orientation—which is to be expected, as each successive layer is smaller than the preceding one—strongly accentuated by the weakening effect of the increasing thickness of the crystal. The effect of repeated diffraction from successive rotated layers is shown by figs. 19 (*a*) and (*b*) where it produces asymmetric spikes around a number of diffraction spots. It also produces spikes around the primary beam (just visible in fig. 19 (*b*)). Moirés caused by such spikes consisting of discrete spots are expected to be complex as they would be superpositions of several moiré lattices having the same symmetry but different spacings. A complex multiplicity of the fringes observed particularly in the thicker parts of our crystals might well be due to this cause.

The existence of a rotational displacement of layers is most strikingly demonstrated by the accidental preparation of fig. 15 where this slight rotation is regularly repeated always in the same sense. This is also reflected by the spotted tails following the main diffraction spots in the

diffraction pattern (fig. 16). We are unlikely to have a loosely superimposed array of otherwise independent crystals but rather an inherently twisted lattice. One manifestation of a deformation due to crystals of differing orientations being in contact, would be the existence of a dislocation grid which in view of the preceding arguments is very likely to contribute to the observed moirés in some cases at least. The deformation, however, must be more profound than this. The curving of the moiré fringes seen in most of our photographs (figs. 2, 4,) could only arise if the lattice itself is deformed either by curving or by fanning or possibly by both. The curved moirés are too irregular to permit further analysis of the underlying lattice deformation at the present stage. A particularly characteristic deformation pattern can often be observed around the screw dislocation in the centre of the growth pyramids (fig. 4).

In general, torsional crystal deformations of the kind under discussion have not been considered in crystal growth or deformation studies, because the strain around the axis has been thought to be too large for such lattice imperfections to occur. Preliminary estimates, however, indicate that the strain even in the immediate neighbourhood of the axis of torsion need not be prohibitive for our type of organic crystals. It might be of interest to remark that twisted crystallization with successive crystals continuously rotated with respect to each other has been frequently reported in the older literature in connection with spherulitic growth of a wide variety of substances (e.g. Wallerant 1907). It might be tempting to connect the periodic extinction effects in polyethylene spherulites (Keller 1955) with the mode of crystallization shown by fig. 15. However, such a connection, if it exists, would not be a direct one as in the present case the axis of torsion is the *c* while in the spherulites it is the *b* direction. In crystals as in fig. 15 the angle between successive layers is 2–5°. This visual determination is in rough agreement with the spacing of spots in the diffraction patterns.

If the crystal is exactly perpendicular to the beam all strong reflections will contribute to the diffraction pattern with still appreciable intensity (fig. 22). It has been deduced above that two reflections are sufficient for reproducing the lattice points. Consequently if the crystal is exactly perpendicular to the beam the moiré should give a well resolved pseudo-image of the lattice in the *c* projection. We do find this in parts (figs. 3 and 4). More often, however, only fringes are visible indicating that the particular portion of the crystal is tilted and conditions for diffraction are like those represented in figs. 24 and 25, permitting only one strong reflection with or without its symmetrical pair.

Wherever the lattice itself is resolved we can clearly recognize the pseudohexagonal pattern representing the structure in *c* projection. It is not *a priori* evident whether the dark or white parts represent the molecules as each in itself produces a hexagonal pattern (this is most obvious when inspecting the photographic negative as well, where contrast is reversed). In one isolated case where the resolution was good enough it could definitely be decided that the dark pattern on the print, i.e. electron

deficiency, corresponds to the molecules themselves, as the same localities appeared as discontinuous dark or light dots in print and negative respectively. The lattice points are mostly connected by streaks which, as far as could be judged, were following the [110] directions as referred to the moiré lattice, in agreement with fig. 26 where such streaks resulted from the excess intensity of the primary beam. Again, in general it cannot be decided whether such streaks connect white or dark regions.

As the resolved moiré lattice is always somewhat distorted it is not possible to determine its orientation with respect to the crystal lattice with any reliability. However this orientation could be more clearly defined in the case of unidirectional streaks. The spacing of the fringes is greatly variable and so is their direction. One characteristic feature, the broad striations along [010], has already been discussed. A second frequently observed systematic feature is a striation perpendicular to a {110} face, the fringes often being direct continuations of the streaks in the cross grating connecting the lattice points in better resolved parts of the moirés (fig. 5). Often it was noted that a particular [110] direction is selected by a particular quadrant of the crystal. There are two cases: (1) Denoting the four faces of the lozenge as 110, $1\bar{1}0$, $\bar{1}10$, and $\bar{1}\bar{1}0$, then the striations perpendicular to 110 (i.e. also to $\bar{1}\bar{1}0$) are found mostly in quadrants defined by faces 110 and $\bar{1}\bar{1}0$ (figs. 5, 17, 2 and 12). In other words the moiré streaks are perpendicular to the faces of not their own but of adjacent quadrants. (2) The fringes are perpendicular to the {110} crystal face of the same quadrant in which they are situated. This second type of moiré was only seldom observed in two adjacent quadrants simultaneously. (It is seen partly in fig. 3.)

Interpreted as diffraction moirés the fact that the striations are perpendicular to certain sets of planes means that they arose through double diffraction from these same planes in crystals which are rotationally displaced with respect to each other. Further, the observed mode of selection of only one set of streaks by a given quadrant means that in one particular quadrant only one kind of {110} plane is in reflecting position. It follows that the reciprocal lattice planes are differently tilted with respect to the beam in different quadrants. In order to account for the particular selections of reflections described above either the orthorhombic crystal itself could be kinked along the lozenge diagonals, the crystal as a whole being in an oblique position with respect to the beam, or the crystals could be non-orthorhombic (e.g. monoclinic), twinned at the diagonals. In the case of the first alternative the kinking and tilting of the orthorhombic crystal would have to satisfy special conditions simultaneously, the probability of which is slight even if there are signs of discontinuities within the crystals along the lozenge diameters (an observation which is also consistent with the second explanation) and some random inclination of the crystals as a whole is possible due to the lack of final control in our setting within the electron microscope. In the case of the second alternative of a twinned non-orthorhombic structure effects (1) and (2) could arise from the same obliquity (deviation from orthorhombic symmetry),

the two different directions of inclination being associated with different type quadrants in the two cases.

The obliquity in the lattice indicates a shear within certain localities of the crystal, having similar magnitudes but different directions in neighbouring quadrants. As will be discussed later there is evidence of some buckling or crumpling within the layers (not a buckling of the crystal as a whole) which might well be associated with the phenomena under discussion.

Deviations from orthorhombic symmetry due to selective shear, probably associated with some internal buckling, shows up most clearly through the displacement of white ghost images of the broad Bragg extinction fringes (fig. 9) and of prominent outlines of the crystals. The Bragg fringes follow the crystal faces but are not quite parallel to them. As mentioned earlier these white ghosts are dark-ground images due to a principal diffraction admitted by the objective aperture, displaced by the spherical aberration at the outer parts of the lens, the displacement being normal to the planes producing the reflection. It is seen from the direction of the displacement that these extinction lines are due to {110} reflections. In a particular quadrant the displacement is normal to the face bounding the quadrant. This means that the extinctions are produced by different {110} reflections in each quadrant, the indices of the reflection and bounding face of the quadrant being the same in the cases observed. Thus again a selection of reflections occurs which requires a bend in a tilted crystal or more likely a local deviation from orthorhombic symmetry. The above selection in particular is consistent with the moirés of case (2). We found no clearly defined ghost image displacement corresponding to the other, more frequently observed selection of moiré fringes (case (1)) where reflection and corresponding quadrant have indices of opposite sign. As in this case the shift of the extinction fringes would be more along the fringe direction than transverse to it the ghost images would not be expected to be as readily observable. We observed many further effects concerning ghost images including also ghosts of moiré patterns but these are too complex and not yet sufficiently studied to merit discussion at this place.

It has been pointed out that the obliquities just discussed are very small. They are just sufficient to rotate the reciprocal lattice points, which are already in a position to reflect to some small extent at least, so as to bring them into strongest reflecting position or to take them away from it. Thus within the accuracy of 1° or so the molecules can still be considered as lying perpendicular to the crystal layers.

It is known that dislocations can be shown up directly through double diffraction moirés (Hirsch *et al.* 1956). There are some parts in our photographs where a sharp fringe can be seen to end in the interior of the crystal. Figure 7 is an example. However the moiré fringe is produced, this implies a dislocation with Burgers vector parallel to the interface joining the screw dislocation at the centre.

3.2. *Morphological Features*

In the present stage of the work we cannot do much more than to enumerate the new observations and make some tentative comments. Some of the morphological observations are very puzzling and indicate an entirely new and unexpected range of phenomena, the interpretation of which would be probably essential for the understanding of crystallization in polymers. The present work merely indicates these new possibilities and also the further capabilities of the improved electron microscope technique adopted by us.

Effects like those in fig. 10 occur even in the first layer of the crystal, hence cannot be due to a superposition effect. In the first place the coarse striation appears to divide the crystal in four quadrants. The existence of such distinguishable quadrants is consistent with our ideas about chain folding. It follows from the lozenge-shaped crystal habit, showing only {110} faces, that the plane of the fold is likely to be near a {110} plane. Folding along {110} would lead to four structurally different quadrants with the fold lying in the 110, $\bar{1}10$, $\bar{1}\bar{1}0$ and $1\bar{1}0$ planes (pairs of these being identical representing diametrically opposite quadrants). We do not understand the nature of the quadrants shown by fig. 10, nevertheless the existence of such a subdivision of the crystal is at least in line with our predictions.

As stated earlier the observable coarse striation in fig. 10 is partly due to Bragg extinction, but there is sufficient evidence to indicate that the lines in question are at least partly caused by protrusions, i.e. by genuine morphological features. The extinction itself must be due to a change in orientation of the unit cell. If referred to the axes of the crystal as a whole this altered orientation would represent a departure from orthorhombic symmetry in the particular locality. This agrees with the conclusions reached previously when interpreting the selective diffraction effects. However, this disturbance of the symmetry would also be the consequence of a morphological irregularity of the crystal as for example, a buckling or crumpling of the layers in localized areas along the lines in question, which is made probable by photographs as figs. 11 and 14, and by the shadowed specimen referred to earlier.

We find that these structures are all transient and disappear gradually after the crystal is dried down through a gradual coalescing and broadening of the streaks (figs. 10, 13). In the latter stages all indications of a morphological disturbance disappear and only Bragg fringes remain noticeable, which also broaden and finally disappear. We propose a tentative explanation which might follow from our picture of chain folding. If the fold period changed as the crystal dried down the surface would become corrugated in the intermediate stages of this transition. If the fold lengthened the crystal would have to contract laterally in order to accommodate a reduced number of segments within a given layer. Correspondingly a shortening of the fold would cause an expansion. If this occurred locally it would result in a localized distortion of the crystal face

as observed. Even after the obvious corrugations had disappeared the crystal could still be locally buckled owing to such changes. This would account for the Bragg fringes in figs. 13 and 9, which accordingly would be expected to lie in the same direction as the preceding corrugation as observed. Accordingly the gradual broadening and final disappearance of these fringes would result from the flattening of the crystal. This localized buckling might be responsible for the particular selection of {110} reflections discussed above.

Our low angle x-ray experiments indicate a spacing of 120 Å, identified with the fold period (Keller and O'Connor 1957). (According to latest unpublished results this period varies between 90° and 140° with the temperature of crystallization.) However, in most photographs we find traces of a reflection corresponding to a spacing larger than the above range (200 Å and above), the exact spacing and the relative intensity of which being variable. This in principle at any rate is consistent with the idea of different fold periods, being present simultaneously in varying proportions.

One of the features requiring explanation is the direction of the corrugations and that of the corresponding Bragg extinction fringes. Although the fringes roughly follow the crystal faces they are not exactly parallel to them. The angles bisected by the long and short diagonals of the crystal (*a* and *b* axes) are about 100° and 80° respectively. Even if variations do occur the angle about *a* remains always obtuse and the one about *b* always acute, which is just the opposite to what we find with the angles between the crystal faces, where in case of ideal development the angle about *a* is $66\frac{1}{2}°$ and that about *b* $113\frac{1}{2}°$. Thus the striations in question are not parallel to {110} planes but are inclined to them by about 16°. Such a direction does not correspond to any crystallographically defined plane.

It can be seen from fig. 11 that one protrusion also follows the *b* axis, i.e. the long diagonal of the lozenge formed by the usual protrusions. Here this feature is only along a short length and does not continue till the centre. In some of our other photographs such protrusions along *b* are more pronounced and pass through the centre of the crystal. In one case this feature is preserved even after the rest of the striations have disappeared. It might be noteworthy to recall that protrusions along the *b* axis have been observed by one of us previously in photographs which were taken without the present precautions, and they offered a promising approach for explaining the formation of fibrillar and subsequent spherulitic growth (Keller 1957).

Directions which are not defined crystallographically also occur in the unusual crystals of fig. 15. Here the shape of subsequent layers changes gradually with the angle of rotation until the originally obtuse and acute angles are interchanged, which is similar to the shapes enclosed by the striations just discussed. At the present stage we have no explanation to offer for this curious phenomenon.

Photographs showing the coarse striation often also reveal a fault-line or rather band parallel to [010] (fig. 12). Where these bands end a notch

can be seen at the bounding face of the crystal. In one place at least it can be clearly seen (we have many further examples in other photographs) that the notch is due to a re-entrant {110} plane, the sign of which is the opposite to that forming the main part of the crystal boundary. Thus the crystal consists of repeated (010) twins, each successive twin being smaller than the preceding one as drawn schematically in fig. 27. Accordingly the appearance of a fault is due to the striation changing direction when passing through a region corresponding to the twin which gives rise to the re-entrant face on the boundary. Formation of whole needles by a regular repetition of re-entrant faces of equal size about the *b* direction as symmetry axis is known to occur in paraffins (Tanaka *et al.* 1928). As in this case the structure is truly orthorhombic this is not twinning but rather incipient dendrite formation. In the above case of polyethylene we can speak of twinning because the structure is not truly orthorhombic. The orientation of the lattice, taken in the sense as so to include the fold, changes across the twin boundary, as is also indicated by the change in the direction of the striation.

This mode of twinning or incipient dendrite formation, whichever the particular case might be, can account for a large number of shapes observed in the course of our studies. It can be seen that the existence of re-entrant faces makes the overall shape of the lozenge more elongated than the one defined by the structure. An equal sequence of the two kinds of faces leads to needles. All intermediate stages can be realized by varying the sequence. A gradual change in the relative sizes of the two kinds of faces leads to curving overall outlines. If the faces cannot be individually resolved such crystals might appear as being bounded by genuinely curved faces as has sometimes been observed in the course of our studies. We have observed that the same kind of repeated twinning or incipient dendrite forming process can also occur about the *a* direction as an axis of symmetry. This leads to a distortion of the lozenge along the original short diagonal (*b* axis). Combination of this type of growth along both directions can lead to the appearance of truly dendritic shapes as shown by earlier studies (Keller 1957, fig. 6).

References

Bunn, C. W., 1939, *Trans. Faraday Soc.*, **35,** 482.

Fischer, E. W., 1957, *Z. Naturf.*, **12a,** 753.

Glauber, R., and Schomaker, V., 1953, *Phys. Rev.*, **89,** 667.

Haine, M. E., 1957, *J. sci. Instrum.*, **34,** 9.

Hirsch, P. B., Horne, R. W., and Whelan, M. J., 1956, *Phil. Mag.*, **1,** 677.

Keller, A., 1955, *J. Polymer Sci.*, **17,** 351; 1957, *Phil. Mag.*, **2,** 1171; 1958, *J. Polymer Sci.* (in the press).

Keller, A., and O'Connor, A., 1957, *Nature, Lond.*, **180,** 1289.

Pashley, D. W., Menter, J. W., and Bassett, G. A., 1957, *Nature, Lond.*, **179,** 752.

Tanaka, Y., Kobayashi, R., and Ohno, S., 1928, *J. Fac. Engng Tokio*, **17,** 275.

Till, P. H., 1957, *J. Polymer Sci.*, **24,** 301.

Wallerant, F., 1907, *Bull. Soc. franç. Minér.*, **30,** 43.

Wilman, H., 1951, *Proc. phys. Soc. Lond.* A, **64,** 329.

Quench Hardening in Sodium Chloride Crystals†

By B. H. Kear‡ and P. L. Pratt
Department of Physical Metallurgy, University of Birmingham

[Received July 30, 1958]

Abstract

The quench-hardening of sodium chloride has been studied, optically and mechanically, after various quenching treatments. These experiments suggest that, in addition to quenching strains, the hardening is due to the pinning of dislocations by impurities and their associated vacancies retained in solid solution, and, after fast quenches from close to the melting-point, to the additional difficulties of moving dislocations through a fine dispersion of vacancy clusters. Further evidence for these two hardening mechanisms is obtained from recovery experiments.

§ 1. Introduction

An interesting quenching effect, termed 'quench hardening', has been reported by Maddin and Cottrell (1955). The flow stress of annealed and furnace-cooled single crystals of aluminium was compared with that of similar crystals quenched in iced brine from 600°c. Quenched crystals were found to be much harder than furnace-cooled crystals. Making allowances for quenching strains and other factors, the authors concluded that part, at least, of the quench hardening was due to the condensation of quenched-in thermal vacancies on dislocation lines to form jogs. The work of Levy and Metzger (1955) on internal friction in quenched aluminium appears to confirm this view.

The present work outlines an examination of quench hardening on sodium chloride crystals. As above, quenching strains and quenched-in defects are considered as sources of hardening, and by suitable choice of experiments it is shown that the contribution to the hardening from these two sources can be separated.

§ 2. Experimental Procedure

Specimens were obtained by cleavage from a large melt-grown single crystal of sodium chloride. After cleavage all crystals were annealed in air at 650°c and furnace cooled. The dimensions of the crystals and the details of the quenching treatments depended upon the experiment, and these will be given in the appropriate places in the text.

Quenching strains were examined using the polarizing microscope, and a standard photographic procedure was adopted when the birefringence

† Communicated by the Authors.

‡ Now at T.I. Research Laboratories, Hinxton Hall, Saffron Walden.

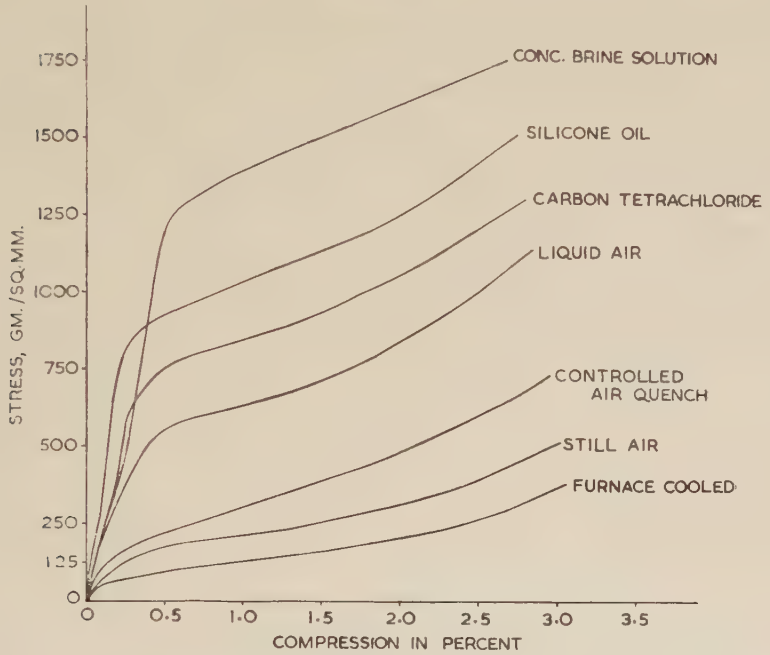
patterns of different crystals were being compared. Strain–stress curves were determined in compression using a ' soft ' machine designed so as to give axial loading.

§ 3. Experimental Results

3.1. *Influence of Quenching Rates*

The first experiments were designed to compare the optical and mechanical effects developed in crystals $3 \times 3 \times 10$ mm by quenching at various rates from 750°c. A variety of quenching media, ranging from

Fig. 2



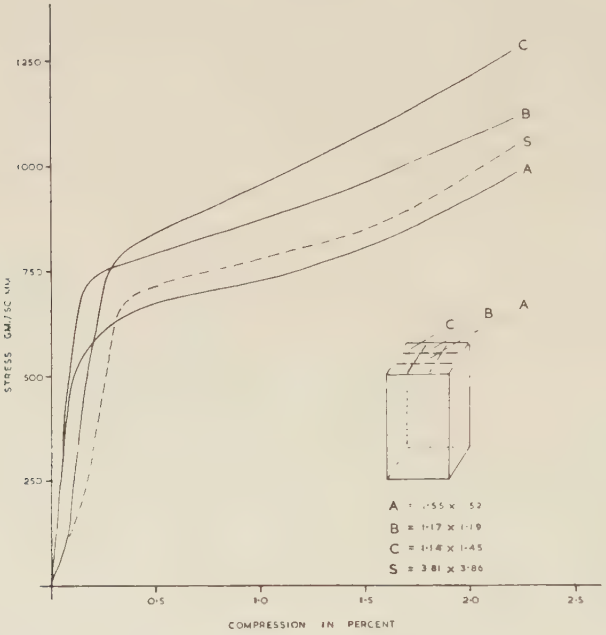
Stress–strain curves for variously quenched crystals.

still air to concentrated brine solution, were used to give the desired variations in the quenching rate. Figure 1, Pl. 17, shows the quenching strain patterns seen through the end and side faces of these variously quenched crystals, and fig. 2 shows the corresponding stress–strain curves. It is clear from these two figures that increasing the quenching rate (i) produces more distortion in the crystal, i.e. increases the severity of quenching stresses and strains, and (ii) markedly increases the quench hardening.

In view of the fact that both quench hardening and the severity of quenching strains increases with the quenching rate, it could be argued that the hardening be due entirely to these quenching strains. However, a careful examination of the results summarized in figs. 1 and 2 gives

support to the view that quenching strains can account only for part of the increase in hardening in the most severely quenched crystals. There is not much difference in the severity of the quenching stresses and strains for the two crystals quenched in silicone oil and concentrated brine solution, but the corresponding stress–strain curves of fig. 2 show a significant difference in the flow stresses of these crystals. This difference is probably due not to quenching strains but to some other cause, which is particularly sensitive to the faster rates of quenching. Figure 2 shows that the flow stress of the crystal quenched in brine is about 1250 g/mm^2,

Fig. 3



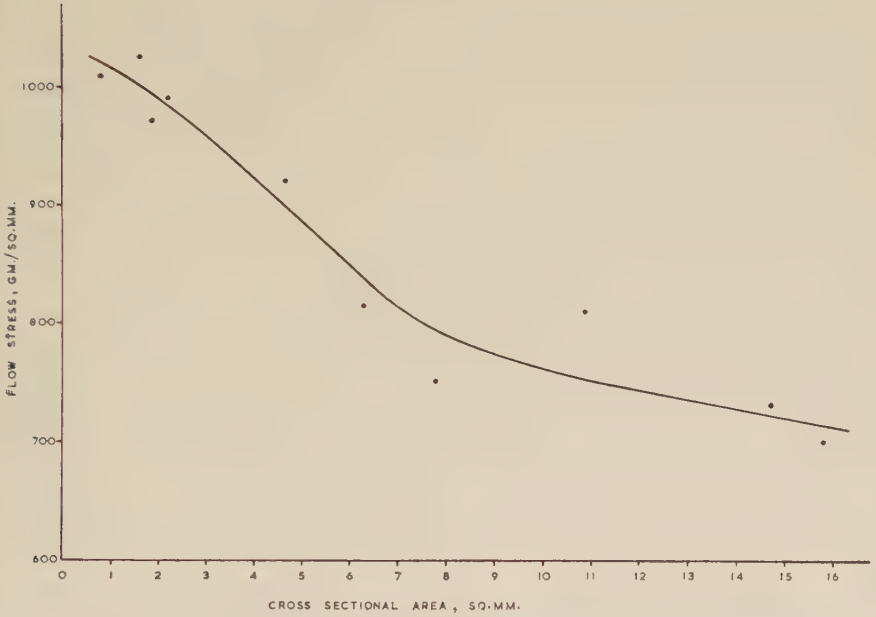
Stress–strain curves of specimens cleaved from different parts of a quenched crystal.

and this hardening is much greater than would be expected from the small plastic strains introduced by quenching. On the other hand these strains do involve slip on a number of intersecting slip systems, as we have shown previously (Kear and Pratt 1958), so that subsequent deformation (to measure a flow stress) necessarily involves dislocation intersections.

Further evidence for the minor effects of quenching stresses and strains was provided by compression tests carried out on different sections of a quenched crystal. Two crystals $4 \times 4 \times 10$ mm were quenched from 750 °C in silicone oil. One of the crystals was cleaved longitudinally to obtain the three smaller crystals A, B and C, as shown in fig. 3, and the other was used as an uncleaved standard S. The planes of cleavage were chosen

to coincide approximately with the extinction planes observed photoelastically, so that the longitudinal residual stresses, originally quite large, were reduced to a very small value. The stress–strain curves for the three sections A, B and C, show that the interior of the crystal is nearly as hard as the surface layers, in spite of the difference in quenching strains. Furthermore, the fact that the flow stress of the standard crystal S is about an average of the values for the three individual sections, shows that longitudinal residual stresses do not exert much influence on the plastic behaviour of the bulk of the crystal.

Fig. 4



Variation of flow stress with cross-sectional area of quenched crystals.

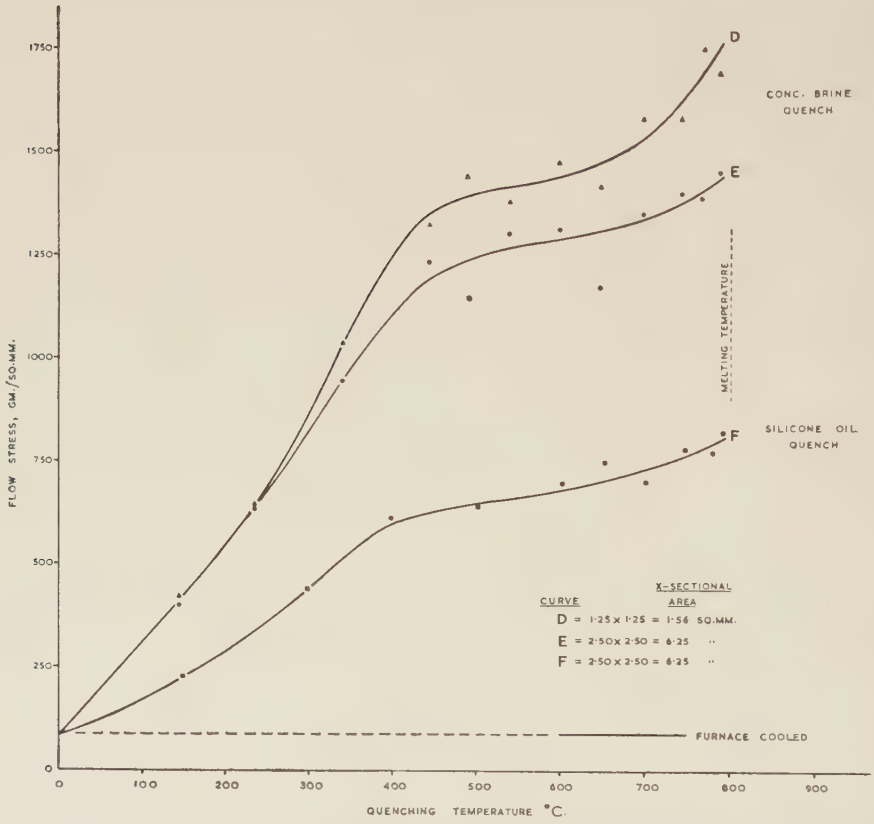
The influence of the quenching rate may also be studied conveniently by varying the cross-sectional area of the crystals being quenched. The results of such an experiment for a batch of crystals quenched in silicone oil from 750°C are shown in fig. 4. The flow stress increases with decreasing cross-sectional area, whereas the severity of the quenching stresses and strains, observed photoelastically, diminished with decreasing cross-sectional area. Obviously this increase in hardening with decreasing cross-sectional area cannot be due to quenching strains.

The conclusions to be drawn from these experiments with different quenching rates may be summarized as follows: quench-hardening is due (i) to the presence of quenching strains, and (ii) to effects arising probably from the quenching-in of point defects, the balance between the two depending sensitively upon the quenching rate.

3.2. *Influence of Quenching Temperature*

Crystals were quenched from elevated temperatures in the range 100–800°c and their flow stresses were measured at room temperature. A plot of the quench-hardening as a function of quenching temperature is shown in fig. 5. Curves D and E compare the hardening of crystals having different cross-sectional areas quenched in the same medium (concentrated

Fig. 5

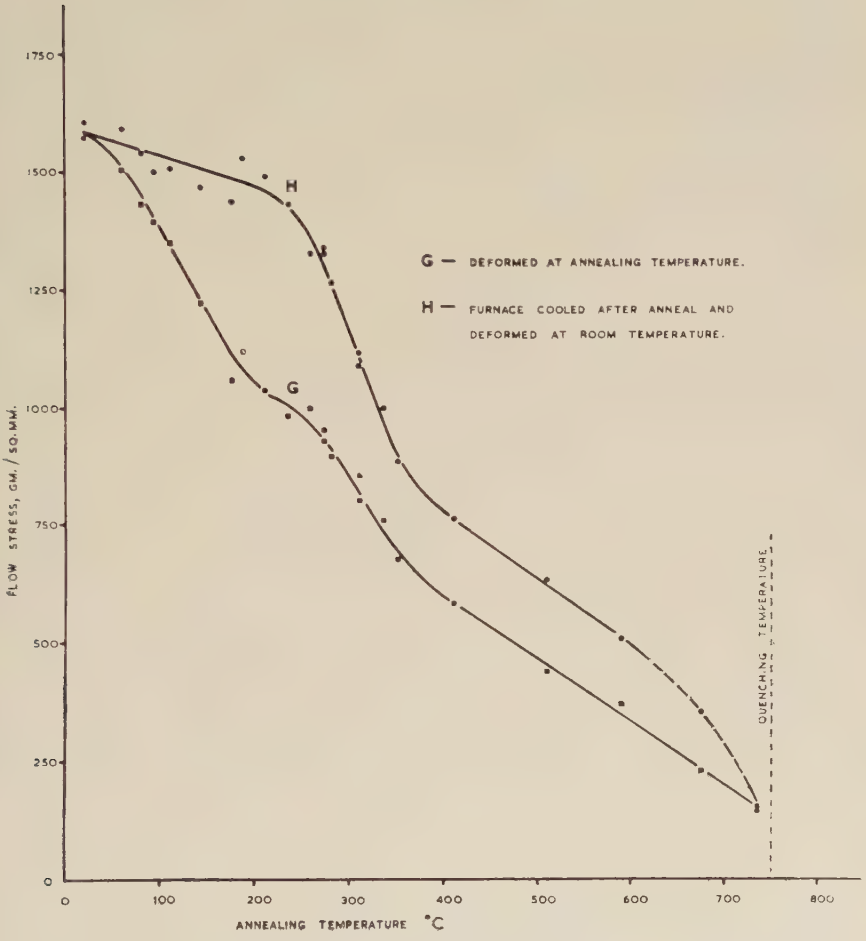


Quench hardening as a function of quenching temperatures for various crystals.

brine solution), while curves E and F compare the effect of different quenching media for crystals having similar cross sections. All three curves show first a rapid increase in hardening with quenching temperature up to about 400°c, followed by a second stage of less marked hardening. Curve D shows a distinct third stage in which the hardening again increases rapidly in the temperature range from 600°c up to the melting point.

These curves also demonstrate how sensitive the hardening is to quenching rate over the whole range of quenching temperatures, the divergence of curves E and F with increasing temperature being particularly striking. Curves D and E show that the effectiveness of a quench in brine in producing hardening is increased by decreasing the cross-sectional area of the crystals in the temperature range from 275°C up to the melting point, and that this effect is more pronounced at temperatures exceeding 600°C.

Fig. 6 (*a*)



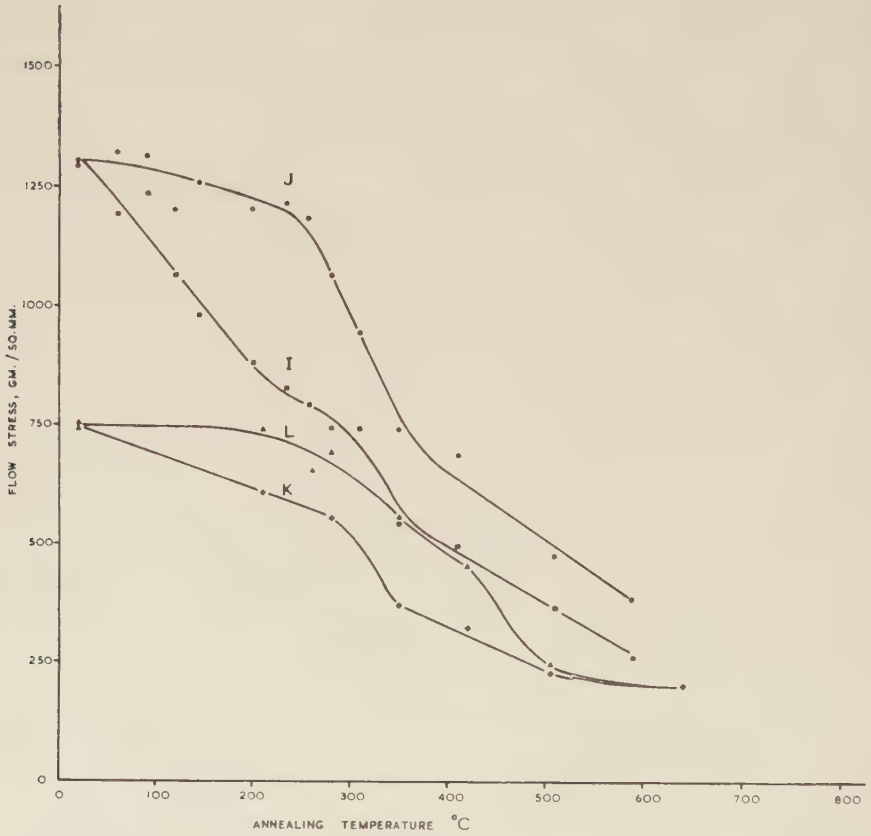
Recovery of flow stress after annealing.

3.3. *Recovery of the Flow Stress after Annealing*

These experiments involved plastic deformation both at room temperature and at elevated temperatures, of crystals that had been annealed after quenching. The procedure was as follows: two quenched crystals were annealed at the selected temperature for 6 hours; one

crystal was deformed at this temperature and the other allowed to cool in the furnace and then deformed at room temperature. Stress–strain curves were recorded in both cases and from these the flow stresses were estimated. This was repeated for a series of annealing temperatures from room temperature up to about 750°C.

Fig. 6 (*b*)



Recovery of flow stress after annealing.

The results of these experiments for three different batches of crystals are shown in figs. 6 (*a*) and (*b*). Figure 6 (*a*) shows annealing curves G–H for crystals ($1{\cdot}25 \times 1{\cdot}25$ mm section) hardened by quenching from 750°C in concentrated brine solution. Figure 6 (*b*) shows two sets of annealing curves I–J and K–L for crystals ($2{\cdot}5 \times 2{\cdot}5$ mm section) hardened by quenching from 750°C in concentrated brine solution and in silicone oil respectively. These variations of crystal section and quenching media were used in the previous experiments, so that these three sets of curves G–H, I–J and K–L may be correlated with curves D, E and F, in fig. 5.

The influence of annealing on the room temperature flow stress of quench hardened crystals is shown by curve H, and on the corresponding flow stress at the annealing temperature by curve G. Curve H shows that softening by annealing occurs in three main stages, in the approximate temperature ranges 20–225°c, 225–375°c and 375–750°c. Between 20°c and 225°c a slight softening is experienced after annealing, and in the second stage (225–375°c) a very pronounced softening occurs, such that at the upper limit of temperature nearly 50% of the original hardening is lost. The remainder of the hardening is lost more gradually in the third stage (375–750°c). It is noteworthy that even at 600°c the flow stress is still about six times that of a furnace cooled crystal. Curve G shows that the flow stress at the annealing temperature decreases steadily with increasing temperature of anneal over the whole range; the decrease in flow stress is quite sharp between 20°c and 375°c but falls off more gradually at temperatures exceeding 375°c. The sharp fall in flow stress between 20°c and 225°c shown by curve G contrasts with the small effect shown by curve H; both curves show about the same decline at higher temperatures.

The annealing effects shown by the pair of curves G–H may be conveniently described in terms of $\Delta\sigma_f$, which refers to the numerical difference between the values of the flow stresses at the annealing temperature and at room temperature after annealing. Thus, the two curves show that $\Delta\sigma_f$ first increases to a maximum of 400 g/mm² at 225°c and then decreases to 200 g/mm² at 375°c, remaining virtually constant at this value up to at least 600°c, after which it falls to zero.

Referring to fig. 6 (*b*), curves I–J show much the same features as curves G–H, even to the extent of an increasing $\Delta\sigma_f$ up to a maximum of 400 g/mm² at roughly 225°c. Curves K–L, on the other hand, show a much reduced numerical value of $\Delta\sigma_f$ in this temperature range, not exceeding 150 g/mm²; moreover $\Delta\sigma_f$ apparently decreases to a very small value at temperatures greater than 500°c. These annealing curves are shown only to 600°c since at temperatures higher than this the stress–strain curves were somewhat irregular in the early stages of plastic deformation, and it was not possible to estimate the flow stresses with certainty.

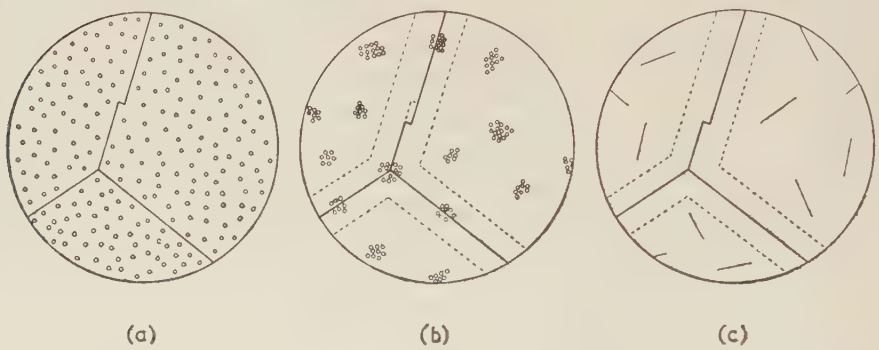
§ 4. Discussion

4.1. *Quench-Hardening of Ideally Pure Crystals*

At temperatures close to the melting point, pure alkali halide crystals contain a large population of thermal vacancies. When these crystals are cooled in such a way as to maintain equilibrium the vacancies are eliminated by condensation in an orderly manner at the free surface and at jogs on dislocation lines. Owing to the high mobility of thermal vacancies this process of condensation should proceed very rapidly, but the final distribution of vacancies should be affected by fast rates of cooling. An infinitely fast quench to a low enough temperature would

retain all the vacancies at rest in the positions originally occupied at the high temperature before quenching; in this case the vacancies will be 'frozen-in'. Such a super-saturated state is comparable with that obtained in precipitation-hardening alloys in which a solution-treatment is followed by a quench. In these alloys raising the temperature permits diffusion of the solute to occur with the formation of clusters and finally precipitates. By analogy, raising the temperature in pure ionic crystals should permit positive-ion and negative-ion vacancies to associate, forming pairs and later clusters. In the vicinity of existing dislocation lines the excess vacancies can disperse in other ways. Probably the formation of dilute Maxwellian atmospheres of vacancies is only the first stage in a process which ultimately produces small groups of vacancies distributed along the dislocations as Friedel (1956) has suggested; in this process of regrouping a proportion of the vacancies may also be lost on

Fig. 7



The clustering of vacancies during annealing after fast quenching.

jogs in the dislocation lines. These changes are illustrated diagrammatically in fig. 7, which compares (*a*), the initial and (*b*) the final states following regrouping of the vacancies. The figure for the final stage emphasizes the zones of denudation of vacancies around the three dislocation lines forming part of the network.

In any real quench, however fast, a certain amount of annealing out of thermal vacancies must occur over the whole temperature range. A proportion of the vacancies will be lost at jogs on the dislocations, and clustering of vacancies may occur at random in the crystal lattice and also along the dislocations. Only in the faster quenches will the result approach the distribution of clusters in fig. 7 (*b*), whereas slower quenching rates may permit redistribution of vacancy pairs within the clusters leading to the formation of prismatic dislocations (Seitz 1952), fig. 7 (*c*). This model should give two different sources of hardening; the dislocation lines in the network may be pinned by the jogs and associated

clusters of defects; secondly, the dispersed clusters and prismatic dislocations provide an obstacle to slip. With a concentration of vacancies of 1 in 10^4 close to the melting point (Etzel and Maurer 1950), condensation into clusters containing 100 vacancies gives a distance between clusters of about 100 atomic spacings. This is, at least, in the range where dispersion hardening might be expected.

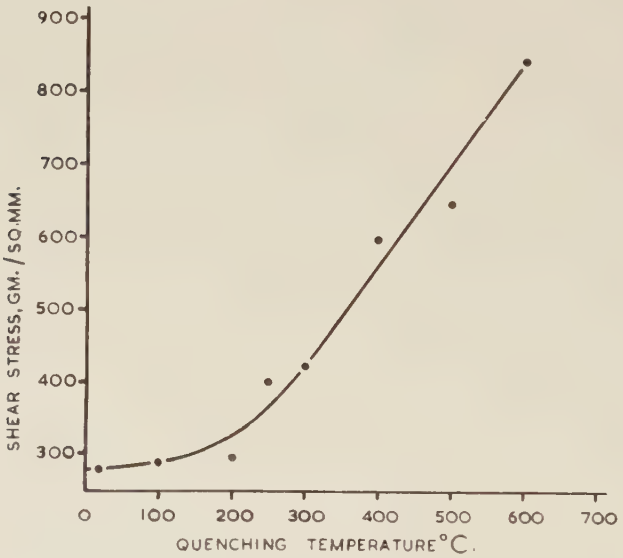
4.2. *Quench-Hardening of Real Crystals*

In addition to thermal vacancies, most ionic crystals contain positive-ion vacancies associated with divalent metallic impurity ions. Following Koch and Wagner (1937), these positive-ion vacancies become dominant at temperatures below the knee of the ionic conductivity curve, and thus are responsible for the enhanced conduction in the impurity range. From our point of view the precise form of the conductivity curve is very important, and, while no measurements were made on the material used for these experiments, crystals grown in as nearly possible the same manner show three distinct regions. At high temperatures the conductivity is intrinsic and due to the equilibrium number of thermal vacancies of both signs, with the positive-ion vacancies carrying the bulk of the charge. Below the knee at about 550°C, a second linear portion of smaller slope is found in which the number of charge carriers is believed to be constant and equal to the concentration of divalent impurity ions, while the slope of the line determines the vacancy mobility. At still lower temperatures, below 350°C, the number of charge carriers decreases as the positive-ion vacancies are removed from solution by association or by precipitation.

Quenching from temperatures below 550°C should retain in solution at room temperature an excess of these impurity ions and their vacancies proportional to the number present at the high temperature. Thus to a first approximation a constant number should be quenched-in from between 550°C and 350°C, and from below 350°C an ever decreasing number. This excess concentration of impurity ions and vacancies should contribute to quench hardening by elastic and electrostatic pinning of the dislocations, and by solid solution hardening. The experimental results in fig. 5 show clearly that the quenching temperature controls the amount of hardening. It is significant that the greater part of the hardening is achieved by quenching from below 400°C where the positive-ion vacancies associated with impurities are predominant. By analogy with the known effects of divalent impurities on the flow stress of ionic crystals, the magnitude of this hardening suggests than an additional impurity concentration of about 1 part in 10^6 is present in solution, and it should prove possible to verify this by independent measurements of conductivity and of dielectric loss after quenching. Little further increase in hardening occurs until the quenching temperature is greater than 600°C, where thermal vacancies begin to play the controlling role.

The rate of quenching from a given temperature influences the degree of hardening in the manner shown in detail by curves D, E and F in fig. 5. From below 400°c slower quenching rates permit a closer approach to equilibrium so that fewer impurity ions and vacancies are available for hardening. From above 600°c the rate of quenching is more important because the size of the clusters of thermal vacancies of both signs will depend on the extent to which association and migration can take place during the quench. Appreciable dispersion hardening from this source is found only in crystal D, which was quenched most rapidly; by comparison crystal E, quenched more slowly owing to its larger cross section shows

Fig. 8



Quench-hardening in aluminium (after Cottrell and Maddin).

only a small contribution. This sensitivity to quenching rate is understandable since the high mobility of vacancies, and the attraction between vacancies of opposite sign, both promote ' over-ageing '.

These results may be compared with those of Maddin and Cottrell (1955), plotted in fig. 8 from table 2 of their paper. In aluminium the quench-hardening increases steadily with temperature, which contrasts with the appearance of the plateau at 400°c in sodium chloride. This difference lends support to the view that hardening of sodium chloride by quenching from below 400°c is due largely to impurity ions and their vacancies.

4.3. *Recovery of Flow Stress*

Summarizing the discussion so far, quench-hardening appears to be due to the following causes:

1. Quenching strains.
2. Impurity-vacancy pinning and solid solution hardening.
3. Dispersion hardening due to vacancy clusters.
4. Collapsed vacancy clusters in the form of prismatic dislocations.

The last two sources of hardening are observed only with fast quenches from close to the melting point.

A better understanding of the nature of these mechanisms may be obtained from the recovery experiments summarized in fig. 6 (*a*) and (*b*). Here two main recovery stages were found, one in the temperature range 225°C to 375°C, and the second from 400°C up to the melting point. Between room temperature and 375°C the flow stress is markedly temperature dependent, a characteristic of the first two sources of hardening listed above, and this implies that the first recovery stage is due to the elimination of these contributions to the hardening. On the other hand above 370°C, $\Delta\sigma_f$ (the difference between the flow stress at room temperature after annealing and that at the temperature of anneal) is small and constant, and this is consistent with a dispersion hardening mechanism. Support for this analysis of the recovery curve comes from a comparison of curves G–H and I–J in fig. 6 with their counterparts D and E in fig. 5. While the general shape of the curves G–H and I–J is similar, G–H is displaced to higher stress levels at all temperatures. Clearly the greater part of this increased hardening only anneals out at the higher temperatures in the second stage of recovery, which suggests that dispersion hardening is more important in curves G–H. This correlates with the increase in hardening noted in curve D for crystals quenched from above 550°C.

Contrasting with curves G–H and I–J, curves K–L show a much reduced temperature dependence of the flow stress at the lower annealing temperatures, a smaller fractional recovery in the first stage, and complete annealing by 500°C. Presumably an even slower quench would lead to the disappearance of impurity-vacancy hardening, and any hardening annealing out below 400°C would have to be accounted for in terms of quenching strains. The birefringence due to the quenching strains was reduced after annealing at 400°C, and eliminated by annealing at 500°C. In agreement with this, Obreimow and Schubnikow (1927), and Newey (1958), show that recovery of the flow stress of work-hardened crystals takes place in this temperature range. From as yet unpublished experiments on intersecting slip we expect that the flow stress of quenched crystals, arising solely from quenching strains, would be about 400 g/mm² if a few per cent of plastic strain occurred during quenching on a number of slip systems.

On curves G, I and K, the flow stress measured at the annealing temperature shows a bump centred at about 300°C. It is interesting to

note that the flow stress of annealed virgin crystals increases from room temperature to a maximum at about 300°c and the size of this increase could account for the bump in the recovery curves. This maximum in the flow stress has been attributed to the existence of a charge on the dislocations (Eshelby *et al.* 1958).

With the identification of the low temperature recovery stage as the elimination of impurity-vacancy hardening and the high temperature stage as the aggregation of vacancies, these recovery mechanisms will now be considered in more detail with reference to curves G–H. Plainly the elimination of impurity vacancy hardening can only occur when the excess positive-ion vacancies become mobile, and Pratt (1957) has deduced that this occurs at about 250°c. This temperature corresponds well with the start of the first stage of recovery in fig. 6 (*a*). The second stage of recovery occurs steadily over the temperature range 400°c to 750°c, in contrast to the abrupt softening which was found at 250°c in the first stage. Two different physical processes occur in this temperature range, and both will contribute to the softening. ' Over-ageing ' involves redistribution of the existing quenched-in thermal vacancies into a coarser distribution of clusters. In addition, the equilibrium number of thermal vacancies increases exponentially with temperature, and above 550°c free vacancies may well be formed by dispersal of the existing clusters. Thus recovery should occur slowly over this whole temperature range with a faster rate in the region of intrinsic conductivity above 550°c.

4.4. *Forms of Quench-Hardening*

The recovery experiments discussed in the previous section confirm the view that at least two distinct hardening mechanisms are required to account for quench-hardening of ionic crystals, in addition to quenching stresses and strains. The form of hardening annealing out at low temperatures is an impurity-vacancy hardening, whereas that annealing out at high temperatures is some form of dispersion hardening. The latter may be due either to dispersed groups of clustered vacancies, or to dispersed prismatic dislocatons formed by the collapse of these vacancy clusters. In metals Cottrell (1957) has criticized the idea that dispersed clusters of defects can account for irradiation-hardening, and he concludes that the hardening is best explained by the defects becoming attached to the centres of dislocations. Kimura *et al.*, private communication, on the other hand, have assumed that vacancy clusters collapse to form prismatic dislocations if more than ten vacancies are in the cluster, and they have analysed in great detail the resistivity-changes found in quenched gold by Bauerle and Koehler (1957). Their analysis indicates that the residual resistivity, after annealing out the mobile vacancies, is due to the stacking faults associated with prismatic dislocations. These dislocations presumably should contribute to quench-hardening in addition to the vacancy pinning mechanism of Maddin and Cottrell,

In considering this dispersion hardening by point defects the important question is whether the vacancies precipitate as spheres, or as platelets which can collapse to form dislocation loops. From theoretical considerations Frank (1957) has concluded that for a small number of vacancies the sphere is more stable, while for a large number of vacancies the collapsed dislocation loop has the lower energy. However, the sphere will have to change into a platelet before the loop can form, and this represents a higher energy state if there are more than about ten vacancies in the sphere. The controlling factor is the surface energy, which is not accurately known, and Frank emphasized that an appeal to experiment is necessary. For ionic crystals these general arguments lead to the formation of spheres, whereas for metals dislocation loops seem more probable. In an effort to decide experimentally between these two possibilities, a few quenched crystals were decorated with gold chloride, and details of these experiments are given in the Appendix. It appears that small cavities are produced by quenching and that the addition of gold chloride is necessary to make them grow to visible dimensions, for no trace of these cavities could be seen in crystals subjected to the same annealing treatment but without gold chloride. The distribution of the cavities makes it clear that they were nucleated at the time of quenching, and not during the subsequent decoration treatment. No trace of decorated prismatic dislocations could be seen with the optical microscope but this may be because cavities are preferred sites for decoration. The observations on wavy slip may be regarded as evidence for nucleation of cavities on screw dislocations. That screw dislocations should be favourable sites for nucleation and growth of cavities seems reasonable, since the case is analogous to that of crystal growth from a supersaturated solution. In the latter case, it is well established that nucleation on screw dislocations can occur at quite low supersaturation.

Acknowledgments

The authors wish to thank Professor G. V. Raynor for providing research facilities in the Department of Physical Metallurgy, and for arranging financial support for one of us (B. H. K.); Mr. A. Taylor for allowing us to make use of his conductivity measurements in the Discussion; and Dr. R. W. Whitworth for invaluable criticism of the manuscript.

APPENDIX

Decoration of Quenched Crystals

By courtesy of Dr. J. W. Mitchell a few quenched crystals were decorated with gold chloride in his laboratory at Bristol University. The technique has been described by Barber *et al.* (1957). In this case the crystals were quenched from 750°c into brine, and annealed with gold

chloride for five hours at 620°C. From our interpretation of the recovery of flow stress curves, this treatment should complete the first stage of recovery, but leave the second stage largely unchanged. Thus any new features found in the decorated crystals should be associated with the high temperature hardening process.

The structure of thin longitudinal and transverse sections, cut from the middle of a quenched crystal, are shown in figs. 9 (*a*) and (*b*), Pl. 18. Both sections show a heavily banded distribution of decorated cavities, lying on surfaces roughly parallel to the external surfaces of the crystal. This layered structure is not found in decorated virgin crystals, nor in quenched crystals that were not decorated. In the latter case no trace of cavities could be found using an ultra microscope. This suggests that the cavities were developed to visible size only by the decoration, but their distribution makes it clear that they had been nucleated during the quench. The origin of the banded structure reflects the changing system of internal stresses during quenching; for nucleation of cavities will be favoured in regions of hydrostatic tension, and inhibited in regions of compression. A particular instance of this influence of the internal stress is shown by the absence of cavities around surface cracks in fig. 10, Pl. 18, which formed during quenching to relieve the tensile stresses. Furthermore, the banded structure deep in the crystal is visibly disturbed by the stress concentration at the tip of these cracks. In the transverse section, fig. 11, Pl. 18, clear evidence of wavy slip could be seen, with cavities revealing the slipped planes. At higher magnifications many more smaller cavities became visible but no trace of the dislocations could be seen. It is clear therefore that the cavities were nucleated at dislocations during quenching and that the latter were subsequently removed by the annealing treatment that produced the decoration. Evidence for plastic deformation during quenching, which we have discussed in our earlier paper (Kear and Pratt 1958), comes from the distortion of the surfaces of the crystal, seen in fig. 9. The banded structure clearly follows this distortion. All of this suggests that the cavities are nucleated in favourable stress zones, and that planes containing such zones move inwards during quenching to account for the banded structure.

References

Barber, D. J., Harvey, K. B., and Mitchell, J. W., 1957, *Phil. Mag.*, **2**, 704.

Bauerle, J. E., and Koehler, J. S., 1957, *Technical Report* No. 8, DA-11-022-ORD-1212.

Cottrell, A. H., 1957, *Symposium on Vacancies and Other Point Defects in Metals and Alloys* (Inst. Metals), p. 1.

Eshelby, J. D., Newey, C. W. A., Pratt, P. L., and Lidiard, A. B., 1958, *Phil. Mag.*, **3**, 75.

Etzel, H. W., and Maurer, R. J., 1950, *J. chem. Phys.*, **18**, 1003.

Frank, F. C., 1957, *Dislocations and Mechanical Properties of Crystals* (New York: John Wiley), p. 514.

Friedel, J., 1956, *Les Dislocations* (Paris: Gauthier-Villars).

KEAR, B. H., and PRATT, P. L., 1958, *Acta Met.*, **6**, No. 7, p. 457.
KOCH, E., and WAGNER, C., 1937, *Z. physikal Chem.* B, **38**, 295.
LEVY, M., and METZGER, M., 1955, *Phil. Mag.*, **46**, 1021.
MADDIN, R., and COTTRELL, A. H., 1955, *Phil. Mag.*, **46**, 735.
NEWEY, C. W. A., 1958, *Thesis*, University of Birmingham.
OBREIMOW, I. W., and SCHUBNIKOW, L. W., 1927, *Z. Phys.*, **41**, 907.
PRATT, P. L., 1957, *Symposium on Vacancies and Other Point Defects in Metals and Alloys* (Inst. Metals), p. 99.
SEITZ, F., 1952, *Advanc. Phys.*, **1**, 43.

Direct Observations of Defects in Quenched Gold†

By J. Silcox and P. B. Hirsch
Crystallographic Laboratory, Cavendish Laboratory, Cambridge

[Received August 21, 1958]

Abstract

An electron microscope transmission study has been made of thin foils of gold produced by electro-polishing from quenched gold specimens. Thin gold sheet was quenched from a vertical tube furnace at 910°c to 960°c into iced brine. Subsequent ageing treatment was carried out, for example for an hour at 100°c. Contrast effects were observed of geometrical shapes with sides parallel to [110] projections in the plane of the foil. Analysis of the shapes of the contrast in two different orientations shows that they arise from tetrahedra of stacking faults on (111) planes with $\frac{1}{2}$[110] type stair-rod dislocations along the edges of the tetrahedron. A reaction is pointed out whereby the defect may arise from an equilateral triangle, with [110] edges, of stacking fault surrounded by a Frank sessile dislocation on a (111) plane. On the assumption that the triangles were formed by the disc-like condensation of vacancies, the size of the defects, about 350 Å, and the density, about 5×10^{14} cm^{-3}, correspond to an initial vacancy concentration of 6×10^{-5}. The interaction of dislocations with the defects confirms the hypothesis of stacking fault tetrahedra. Finally the formation of the defects and their influence upon the physical properties of the crystal are discussed.

§ 1. Introduction

One of the outstanding problems of dislocation theory is the origin of dislocations in metal crystals grown from the melt or in well-annealed crystals. Frank (1950) suggested that dislocation loops might be produced by the collapse of discs of vacancies formed by precipitation of these defects under conditions of supersaturation during cooling of the specimen. This idea was elaborated by Seitz (1950) who suggested that prismatic dislocation loops would be produced in this way, and that these might act as dislocation sources and also, when suitably arranged, constitute the substructure of crystals. The formation of substructures as a result of vacancy condensation has been considered in detail by Teghtsoonian and Chalmers (1951, 1952) and by Frank (1956). Some features of the condensation process have been discussed by Read (1953) and by Barnes (1954). More recently the presence of dislocation loops has been invoked to explain the experiments of Gilman and Johnston (1957) on LiF (Fisher 1957) and the phenomenon of quench-hardening (Kimura *et al.* 1958 a, b, c). The most detailed discussion of the formation and properties of dislocation loops resulting from the collapse of discs of vacancies has been given by Kuhlmann-Wilsdorf (1958). In addition to a number of

† Communicated by the Authors.

predictions this author also drew attention to the difference expected between face-centred cubic metals of different stacking fault energies. In a face-centred cubic metal the collapse of a disc of vacancies on a (111) plane results in the formation of a stacking fault bounded by a Frank partial dislocation with Burgers vector $\frac{1}{3}[111]$. If the stacking fault energy is high, as for aluminium, it is energetically favourable to nucleate a Shockley partial to remove the fault and to combine with the Frank partial to form a total prismatic dislocation (referred to as an R-dislocation by Kuhlmann-Wilsdorf) according to a reaction of the type

$$\tfrac{1}{3}[111]+\tfrac{1}{6}[11\bar{2}]=\tfrac{1}{2}[1\bar{1}0]. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

In metals of low stacking fault energies, the loops are expected to be sessile Frank partials bounding stacking faults, although, as suggested by Kuhlmann-Wilsdorf, some of the loops may be converted into glissile prismatic dislocations by the above reaction as a result of thermal activation or of local stresses due to other dislocations.

Dislocation loops have been observed directly in crystals of NaCl (Amelinckx 1957) and of AgCl (Jones and Mitchell 1958). In the latter case the dislocations are generated by the high stress concentration at impurities: in the former case the origin of the loops is somewhat in doubt, but the nature of some of the loops is consistent with a process of formation by the collapse of discs of vacancies.

Direct evidence for the formation of loops by this process has been reported by Hirsch *et al.* (1958) in specimens of aluminium (a metal of high stacking fault energy) quenched from a high temperature. Electron microscope transmission micrographs of quenched specimens showed the existence of a high concentration of loops several 100Å in diameter. In agreement with the predictions of Kuhlmann-Wilsdorf for a metal of high stacking fault energy, the loops are found to be glissile and of the type expected according to the above reaction (1).

It was of interest to repeat the experiments on a face-centred cubic metal of low stacking fault energy to detect if possible the dislocation loops or stacking faults. This experiment was carried out at the suggestion of Professor R. Maddin, and gold was chosen for this study as electrical resistivity measurements on quenched gold had already been interpreted in terms of stacking faults arising from collapsed discs of vacancies (Kimura *et al.* 1958 a, b, c).

§ 2. Experimental Techniques

Specimens of polycrystalline gold foil, 99·998% pure, 0·075 mm thick, were given a two-hour anneal at about 930°C. They were then quenched from a vertical tube furnace at temperatures of between 910°C to 960°C into iced brine. Finally they were aged for periods of about one hour at temperatures between 100°C and 250°C. Following this treatment, the specimens were thinned in a bath of 17 g potassium cyanide, 3·75 g

potassium ferrocyanide, 3·75 g potassium sodium tartrate, 3·5 cm^3 phosphoric acid, 1 cm^3 ammonia, and 250 cm^3 of water (Murphy–Tomlinson—reported by Whelan 1958a). The specimens were then washed in glass-distilled water and examined in a Siemens Elmiskop I electron microscope operating at 80 kv. For comparison purposes, a specimen was cooled slowly from 960°C over a period of about 5 hr and allowed to furnace cool from there. The foils after this treatment had a [100] preferred orientation.

§ 3. Contrast

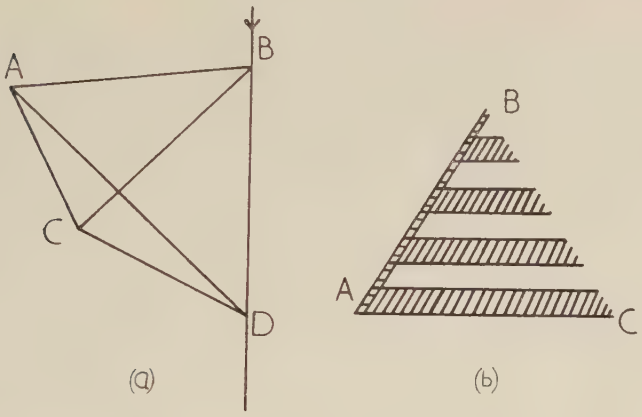
Figures 1–3, 5, 7, 9, 10, 12, 13, 14 (Plates) are micrographs of typical areas in the quenched specimens. They all show contrast effects in the form of geometrical shapes with sides parallel to the projections of the [110] directions on to the plane of the foil. Figures 1 and 2 are low and high magnification micrographs of areas with the normal to the foil approximately parallel to [100], as determined by selective area electron diffraction. It should be noted that the geometrical shapes are squares (A), elongated hexagons (B) and straight lines (C and D). Figure 3 shows another area of a specimen in an orientation close to [110]. In this case, the shapes are triangles (E) and straight lines (F). The possibility that these effects were caused by crystallographic etching has been considered. In this case the contrast would in general be light, relative to the background and fringes parallel to the edges of the pit would be expected. However, the experimental observations show that the contrast is always dark relative to the background and the fringes are of a different kind. Furthermore no trace of these shapes has been observed in slowly cooled specimens (see fig. 4, Pl. 22) for a micrograph of a slowly cooled specimen) while enhancement of contrast near some of the extinction contours shows that the contrast is due to the Bragg diffraction mechanism (Hirsch *et al.* 1956, Whelan *et al.* 1957). Thus the contrast is due to a defect lying within the foil. The simplest and most obvious figure to give rise in projection to the square and triangle shapes shown in figs. 2 and 3 is a regular tetrahedron, consisting of equilateral triangles on (111) planes, the edges of which are parallel to [110] directions. This tetrahedron is in fact identical in shape with Thompson's reference tetrahedron (Thompson 1953).

In most cases, some form of variation of contrast can be seen within the shapes (figs. 1–3, 5). To explain this fact cavities or three-dimensional arrays of stacking faults have been considered. In the case of cavities, the contrast would in general again be light relative to the background while stacking faults produce the reverse situation. Since experimentally the contrast is always found to be dark relative to the background, it presumably arises from stacking faults. In fact the fringes observed are characteristic of overlapping stacking faults on (111) planes; in particular the squares and triangles observed in the [100] and [110] orientations

respectively show fringes which would be expected from regular tetrahedra (referred to above) consisting of stacking faults on intersecting (111) planes.

As an example, consider the contrast shown in fig. 5. The orientation of the normal to the foil in this case is close to, but not exactly, [110] or BD on fig. 6 (*a*). In this orientation the faults on planes ABD and BDC should be nearly parallel to the electron beam; they will therefore only show up as the edges of the triangle, AB and BC (fig. 6 (*b*)). The planes ABC and ADC however will give rise to fringes parallel to their line of intersection AC. Examination of fig. 5 reveals such fringes, the general appearance of which has been redrawn in fig. 6 (*b*). From the electron diffraction photograph it was deduced that the reflection from planes parallel to BCD was strong in the crystal. This suggests that the

Fig. 6



(*a*) A tetrahedron in the [110] orientation: BD is parallel to [110] and to the normal of the foil.
(*b*) The fringes shown in fig. 5, redrawn for the sake of clarity.

orientation of the crystal is such as to tilt the defect away from the [110] orientation by the few degrees sufficient to bring the planes parallel to BCD into the Bragg reflecting position. In this case the stacking fault shear R on the plane BCD (which contributes to the contrast along BC) will be perpendicular to the reciprocal lattice vector $\mathbf{g}$ and the resultant phase factor due to this fault, $\alpha = 2\pi \mathbf{g} \cdot \mathbf{R}$, will be zero. Under these circumstances the stacking fault on the plane BCD will show no contrast along BC (Whelan and Hirsch 1957 a, b). On the other hand the stacking fault on the plane ABD shows contrast along AB. The contrast should therefore consist of fringes parallel to AC, a pronounced line along AB and weak if any contrast along BC; this is exactly as observed on fig. 5 and redrawn on fig. 6 (*b*).

Figure 7 shows an example of the fringe system observed in the [100] orientation. The contrast is in the form of dots; the variation of contrast along the edges of the square proves that these edges correspond to lines inclined to the plane of the foil, in agreement with the fact that they correspond to the [110] directions of the projected tetrahedron of stacking faults. The dotted nature of the contrast is as expected from overlapping non-parallel stacking faults which intersect in lines not parallel to the plane of the foil. A full discussion of the nature of this contrast will be given elsewhere.

So far, only the square and triangle shapes have been discussed in terms of the tetrahedron of faults. The other shapes are thought to arise from the interaction of the tetrahedra with the foil surfaces by mechanisms to be discussed in § 5.2 Before passing on to these points and to the behaviour of defects, a deeper discussion of their nature and of the mechanism by which they are formed will be of value.

§ 4. Interpretation

As the tetrahedra are found in specimens of Au only if they have been quenched, it is thought that they arise from the collapse of aggregates of vacancies, a process similar to that leading to the formation of prismatic dislocation loops in Al (Hirsch *et al.* 1958). The difference between the defects in Au and in Al is attributed to the difference in stacking fault energy, γ, which is probably about 30 ergs/cm^2 for Au (Suzuki and Barrett 1958, Thornton and Hirsch 1958) and probably about 200 ergs/cm^2 for Al (e.g. Seeger 1957). The value of γ for Au is so low that the stacking faults should be stable (Kuhlmann-Wilsdorf 1958, Hirsch *et al.* 1958). Accordingly the collapse of a disc of vacancies on a (111) plane in Au will lead in the first instance to a stacking fault bounded by a Frank partial. However, the Frank sessile dislocation has a relatively large Burgers vector ($\sqrt{\frac{2}{3}}$ that of a normal unextended $\frac{1}{2}[110]$ dislocation), and will dissociate into a low energy stair-rod dislocation (Thompson 1953, 1955, Friedel 1955, Whelan 1958 a, b) and a Shockley partial on an intersecting slip plane according to a reaction of the type

$$\left.\begin{array}{ccccc} \frac{1}{3}[111] & = & \frac{1}{6}[101] & + & \frac{1}{6}[121] \\ (\frac{1}{3}) & & (\frac{1}{18}) & & (\frac{1}{6}) \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

This reaction is energetically favourable as can be seen from the energies of the dislocations (the figures underneath the Burgers vectors of eqn. (2) are proportional to the squares of the Burgers vectors). The stair-rod dislocation $\frac{1}{6}[101]$ lies at the intersection of two intrinsic stacking faults, and the angle between the faults is acute (e.g. Whelan 1958 a, b). It is convenient for the subsequent discussion to use Thompson's notation for the Burgers vectors (Thompson 1953). Figure 8 (*a*) shows the reference tetrahedron ABCD consisting of (111) planes, the edges (AB etc.) of which correspond to the Burgers vectors $\frac{1}{2}[110]$. The mid points of the

faces are labelled α, β, γ, δ; then the Shockley partials are dislocations with Burgers vectors of the type $\mathbf{A\delta}$, etc., the low energy stair-rods have Burgers vectors of the type $\boldsymbol{\alpha\delta}$, and the Frank partials have Burgers vectors of the type $\boldsymbol{\alpha A}$.

Suppose now that the vacancies condense on the (111) plane α, in the form of an equilateral triangle with edges parallel to the [110] directions BC, CD, DB (this assumption will be discussed in § 6). After collapse, the dislocation loop follows these edges and the Burgers vector is $\boldsymbol{\alpha A}$. The formation of the tetrahedron then proceeds in two stages. The first stage consists of three reactions of type (2), and in Thompson's notation these are:

$$\left.\begin{aligned} \boldsymbol{\alpha A} &= \boldsymbol{\alpha\beta} + \boldsymbol{\beta A} \\ \boldsymbol{\alpha A} &= \boldsymbol{\alpha\gamma} + \boldsymbol{\gamma A} \\ \boldsymbol{\alpha A} &= \boldsymbol{\alpha\delta} + \boldsymbol{\delta A}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Each side of the triangle thus dissociates into the appropriate stair-rod and partial dislocations. Figure 8 (*b*) shows a physical picture of this stage of the process, the partials $\boldsymbol{\beta A}$, $\boldsymbol{\gamma A}$, $\boldsymbol{\delta A}$ bowing out on their slip plane as they are repelled by the stair-rods. Taking into account the fact that opposite ends of the loops are of opposite signs (also indicated on fig. 8 (*b*)) it is found that the partials attract each other in pairs to form stair-rods along **DA**, **BA** and **CA** according to the reactions:

$$\left.\begin{aligned} \boldsymbol{\beta A} + \boldsymbol{A\gamma} &= \boldsymbol{\beta\gamma} \\ \boldsymbol{\gamma A} + \boldsymbol{A\delta} &= \boldsymbol{\gamma\delta} \\ \boldsymbol{\delta A} + \boldsymbol{A\beta} &= \boldsymbol{\delta\beta}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

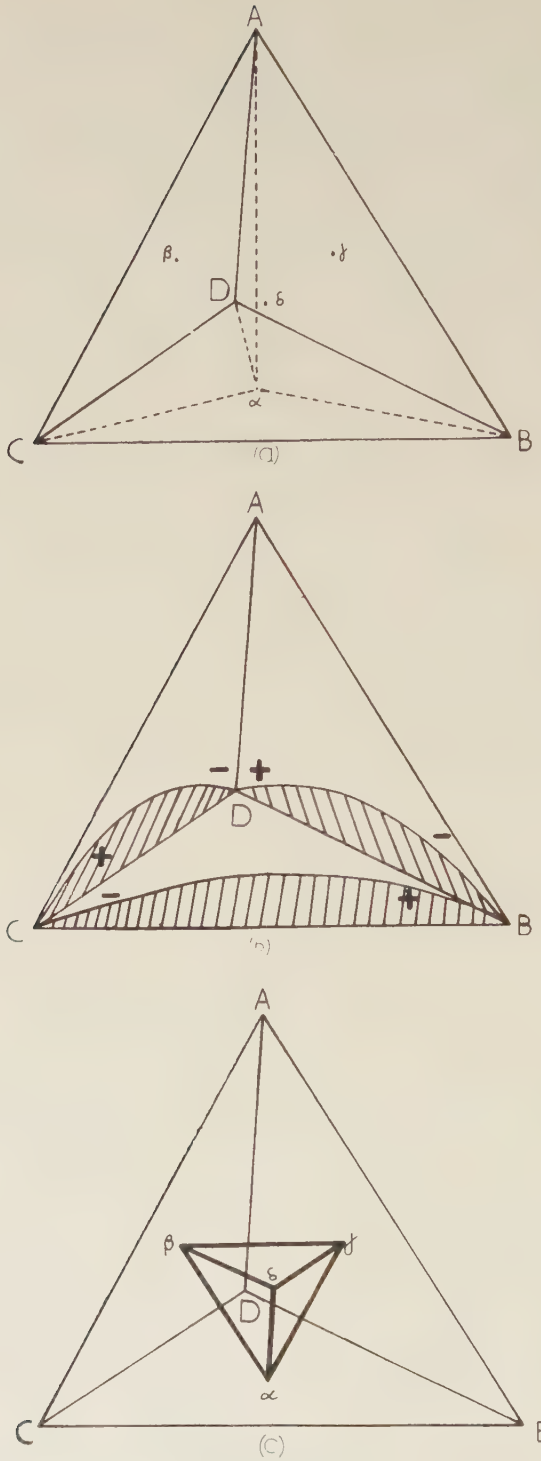
In vector notation the reactions are of the type

$$\left.\begin{aligned} &\tfrac{1}{6}[121] + \tfrac{1}{6}[\bar{1}\bar{1}\bar{2}] = \tfrac{1}{6}[01\bar{1}] \\ &\quad (\tfrac{1}{6}) \qquad (\tfrac{1}{6}) \qquad (\tfrac{1}{18}) \end{aligned}\right\} . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

From the addition of the squares of the Burgers vectors placed underneath the Burgers vectors in (5) it is clear that this reaction is energetically very favourable. The final defect therefore consists of a tetrahedron of stacking faults on (111) planes with the [110] edges of the tetrahedron consisting of low energy stair-rod dislocations of the $\boldsymbol{\delta\alpha}$ type. The Burgers vectors of the stair-rod dislocations also form the edges of another tetrahedron, whose orientation is the inverse of the defect tetrahedron. The relationship between the defect tetrahedron and the tetrahedron of the Burgers vectors of the stair-rods is shown in fig. 8 (*c*). The stair-rods along **AB**, **BC**, **CD**, and **DA** have respectively the Burgers vectors $\boldsymbol{\gamma\delta}$, $\boldsymbol{\delta\alpha}$, $\boldsymbol{\alpha\beta}$ and $\boldsymbol{\beta\gamma}$.

The tetrahedron of stacking faults formed by the above sequence of events is essentially symmetrical. The same final configuration would be attained if collapse had taken place originally on any other (111) plane. In fact the defect may be regarded as the nearest approach possible in a face-centred cubic crystal to the collapse of a spherical shell of vacancies,

Fig. 8



(*a*) Thompson's reference tetrahedron ABCD with the mid-points of the faces labelled, α, β, γ, δ. A Frank partial has a Burgers vector $\boldsymbol{\alpha A}$, a Shockley partial $\boldsymbol{\alpha B}$, $\boldsymbol{\alpha C}$, $\boldsymbol{\alpha D}$, etc., a low energy stair-rod $\boldsymbol{\alpha\delta}$, $\boldsymbol{\alpha\beta}$, $\boldsymbol{\alpha\gamma}$, etc., and an unextended dislocation **AD**, **AB**, **AC** etc.

(*b*) A physical picture of the reaction (3) showing the partials $\boldsymbol{\beta A}$, $\boldsymbol{\gamma A}$, $\boldsymbol{\delta A}$ bowing out in the slip planes ACD, ADB, ABC respectively. The relative signs of the dislocations near the nodes D, B, C, are indicated.

(*c*) The relationship between the defect tetrahedron ABCD and the Burgers vector tetrahedron $\alpha\beta\gamma\delta$.

The energy of the system of stair-rods in the final tetrahedron is proportional to $6 \times \frac{1}{18} = \frac{1}{3}$ compared with $3 \times \frac{1}{3} = 1$ for the original stacking fault triangle bounded by Frank partials. The dissociation therefore leads to a lowering of energy to $\frac{1}{3}$ of the original value, neglecting the energy of the stacking faults. If this energy is taken into account there will be a maximum size for the tetrahedron; a calculation in the Appendix gives the maximum side of the tetrahedron as about 430 A.

§ 5. Results of Observations

5.1. *General*

The lengths of the edges of the tetrahedra are on the average about 350 Å. The largest tetrahedra observed have edges of about 500 Å, in good agreement with the order of magnitude estimate given in the Appendix. The smallest tetrahedra which can definitely be recognized as such have sides about 100 Å long; however, the existence of tetrahedra or of other defects with diameter less than about 100 Å cannot be ruled out. There seems to be little difference in the size or distribution of tetrahedra in specimens given different annealing treatments, but so far the annealing treatments have only been varied over a limited range (see § 2).

In order to determine the density of the defects the thickness of the foil must be known. This may be determined from the width of a slip trace such as that shown in fig. 9 (*b*), provided that the orientation of the foil is known, and assuming that the slip plane is a (111) plane. In this case the thickness of the foil is about 1000 Å and the corresponding average density of defects is about $5 \times 10^{14}\,\mathrm{cm}^{-3}$. The average number of vacancies per defect is about 7400, and the corresponding total vacancy concentration is about 6×10^{-5}. This figure compares well with the values of about 8×10^{-5} and of about 2×10^{-4} deduced respectively by Kimura *et al.* (1958 a, b, c) and by Koehler *et al.* (1957) from the electrical resistivity data of Bauerle and Koehler (1957) on Au quenched from 900°C.

The density of other dislocations observed in these specimens is thought to be slightly higher than in the Al specimens (Hirsch *et al.* 1958), and is about $5 \times 10^{8}\,\mathrm{cm}^{-2}$. This value is to be compared with the density of stair-rods in the defects, about $10^{10}\,\mathrm{cm}^{-2}$, and the density of the corresponding Frank sessile loops is $5 \times 10^{9}\,\mathrm{cm}^{-2}$.

Owing to the large grain size of the specimen and to difficulties in obtaining suitable microscope specimens, it has so far not been possible to obtain good micrographs of grain boundaries.

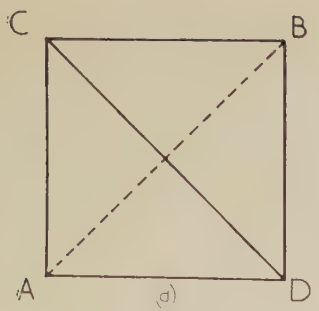
Figure 4 shows a micrograph of a slowly cooled specimen for comparison purposes. To obtain representative pictures of the defects present in the specimen the stereoholder was used to tilt the specimen and to sweep the extinction contours across the specimen. No tetrahedra or similar defects were observed; the density of dislocations was found to be about $5 \times 10^{8}\,\mathrm{cm}^{-2}$.

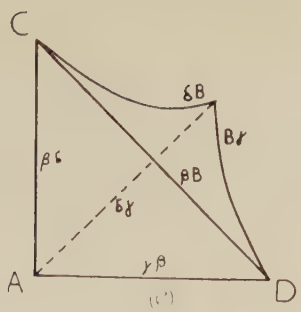
5.2. *Effect of Foil Surface*

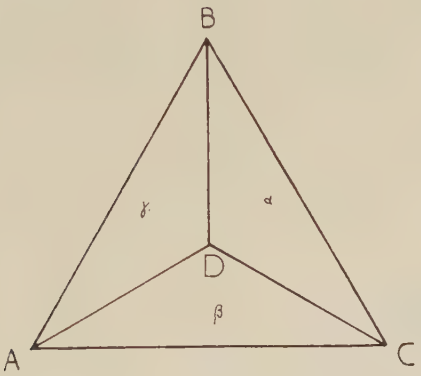
As was mentioned in § 2, apart from tetrahedra other shapes are observed which appear either as elongated hexagons (figs. 1, 2, 9, 10, 12, 13) or as straight lines parallel to [110] directions lying in the plane of the foil. The possibility has been considered that these defects might be other equilibrium shapes arising for example from hexagonal or trapezoidal loops on (111) planes or from collapsed discs on other planes; however, although the observed shapes can be explained on suitable models, such models when viewed from another equivalent crystallographic direction would have a different appearance. This is not observed.

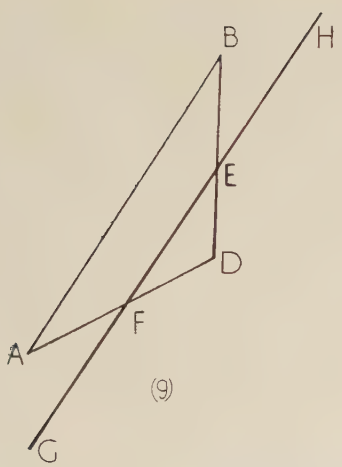
All the observed shapes can, however, be explained easily as a result of intersection of the surface with the tetrahedra. Since the thickness of the foil is only about two or three times the size of the tetrahedra a large fraction of the defects are expected to intersect the surface.

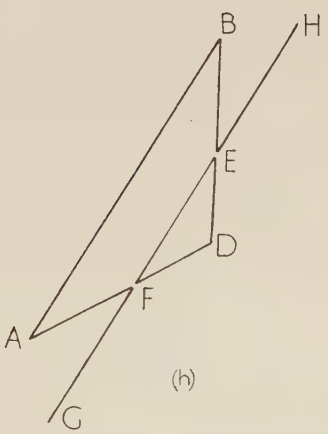
Figures 1, 2, 9, 10, 12, 13, 14, show examples of the foil in the [100] orientation. The elongated hexagons B and the straight lines C, D parallel to [110] directions can be seen clearly. In this orientation two types of tetrahedra are possible, and these are shown in figs. 11 (*a*) and 11 (*b*) respectively. Suppose the surface intersects the defect at EFGH. Then two types of defects which in projection appear as elongated hexagons are produced which are also shown in figs. 11 (*a*, *b*). These shapes are precisely those observed, the elongation being along the [110] directions in the plane of the foil (see figs. 1, 2, 9, 10, 12, 13). The contrast of these shapes confirms in general this interpretation. In the region EFGH the stacking faults are not overlapping and therefore a set of fringes parallel to AB is expected. In the regions AEF and HGB however the contrast should be similar to that observed for the complete tetrahedra, i.e. in the form of dots, as shown schematically in fig. 11 (*c*). The clearest examples of the contrast are shown in fig. 10, and these agree very well with the contrast expected (fig. 11 (*c*)). The lines C, D on figs. 1, 2, 9, 10, 12, 13 are thought to arise in the same way, except that the tetrahedra are cut off very close to lines AB in figs. 11 (*a*, *b*). Some of these lines do actually appear as badly resolved and much elongated hexagons on close examination. The possibility may arise that the defects change their configurations to lower the energy. For example, if AB $\gg$ AF, it is bound to be energetically favourable for the defect to collapse on to the triangle FAE, the Frank partials along AE and AF being split into stair-rods and Shockley partials, the latter however moving only a little way on planes AEHB and AFGB to take up the configuration of minimum energy. However, in order to reach this configuration a Shockley partial must be nucleated, for example on HGB and this requires considerable activation energy. Although this might be possible as a result of high local stresses, in general the defects are expected to remain dissociated, in agreement with observation. Some very small triangles which have been observed, for example at G on figs. 9 (*b*), 10, might possibly correspond to the collapsed configuration, but one cannot be certain of this interpretation.







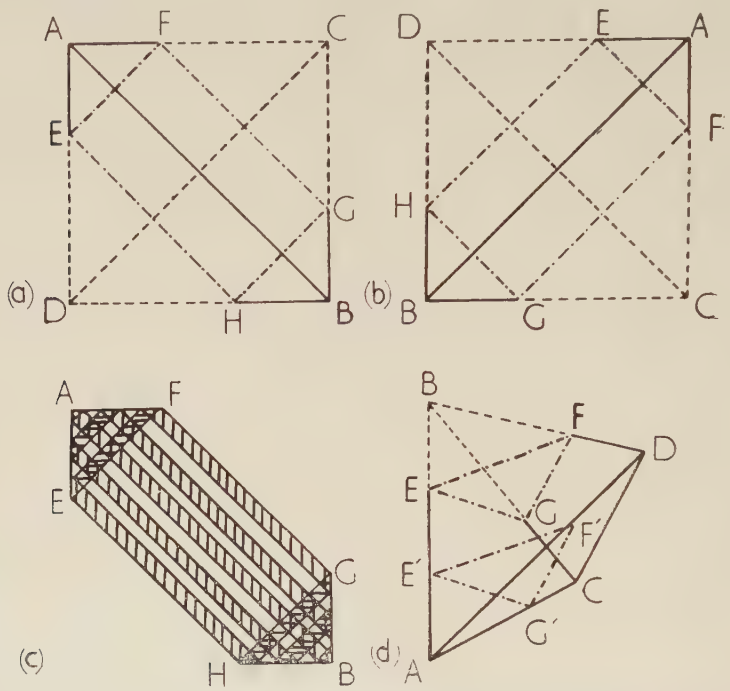




(*d*) Nomenclature used for the tetrahedron J, studied in figs. 10 (*a*, *b*, *c*).
(*e*) A drawing of the tetrahedron J in the state as observed in fig. 10 (*c*). The Burgers vectors of the respective dislocations are given beside them.
(*f*) A dislocation T with Burgers vector **AC** is approaching the tetrahedron ABCD on a plane parallel to ABC and on the same side of ABC as the apex D.
(*g*) This diagram represents the collapse of the tetrahedron ABCD on the plane ABD after a dislocation GFEH has approached it.
(*h*) This is the final state after the reaction (6) has taken place showing the jogged dislocation GFABEH and the smaller loop DFE.

Figures 3 and 5 show examples of the foil in the [110] orientation. In this case the projection of the truncated tetrahedra will still appear as triangles, and it is interesting to note that in fact most of the defects

Fig. 11



(*a*, *b*) Diagrams showing how elongated hexagons AEHBGF may be formed by the foil surface EFGH cutting the tetrahedron ABCD. The solid lines lie within the foil, the dashed lines lie outside, and the dash-dotted lines represent the foil surface. It should be noted that the elongation of the hexagon is always along the [110] directions, lying in the plane of the foil.

(*c*) A diagram representing the type of contrast expected from the hexagon FGBHEA of fig. 11 (*a*). On the planes HGB and AEF there are overlapping stacking faults on intersecting (111) planes giving dotted contrast, and in the region EFGH there are single stacking faults giving rise to striped contrast.

(*d*) A diagram showing a tetrahedron ABCD in the [110] orientation which has been truncated by the foil surface EFG, above the line CD. The figure AE′G′F′ represents the truncated tetrahedron when the foil surface E′F′G′ lies below the line CD. (BA is parallel to [110] and to the normal of the foil.)

appear as triangles on figs. 3 and 5. Figure 11 (*d*) shows schematically the defect when truncated above and below the [110] edge parallel to the plane of the foil. When the defect is truncated above the line CD (fig. 11 (*d*)) order of magnitude calculations indicate that for sufficiently

large values of BE it becomes energetically favourable for the defect to collapse to form the loop FGCD, which would appear as a trapezium projected on to (110). However, this change of configuration would again require considerable activation energy since a Shockley partial must be nucleated. In fact, no such collapsed defect has been observed. When the defect is truncated below CD, for sufficiently large AE′, collapse to form loops AE′F′ or AE′G′ is possible, but again activation energy is necessary. On figs. 3 and 5 some examples can be seen of lines parallel to the directions EF and EG which presumably correspond to Frank sessiles on these planes. It is possible that they were formed by the collapse of tetrahedra.

The above discussion leaves little doubt that most if not all the defects observed can be interpreted in terms of one basic defect in the form of a tetrahedron of stacking faults, and modifications derived from it.

5.3. *Flexibility of Defect*

The tetrahedron of faults is a symmetrical and flexible defect. Under sufficiently high stresses or high temperatures the defect can be made to collapse into a Frank sessile ring on any of the four (111) planes of which it consists, and further nucleation of a Shockley partial can turn the Frank sessile into a glissile prismatic dislocation (Read 1953, Kuhlmann-Wilsdorf 1958). The sequence of micrographs of an area in the [100] orientation shown in fig. 10 illustrates the first of the above two mechanisms. Figure 10 (*a*) shows the two tetrahedra J, K in the complete configuration. In fig. 10 (*b*), however, the tetrahedron J, possibly as a result of thermal stresses due to heating by the electron beam, has collapsed on to the plane ACD (see fig. 10 (*d*))† and the faults on planes ACB. ADB and DCB have disappeared. This reversal of the dissociation can take place by the nucleation of a loop of Shockley partial on plane DCB for example. In fig. 10 (*c*) the faults on planes ACB and ADB are re-emerging beyond the line CD, the front of these faults being bounded by glissile Shockley partials. This can be deduced from the set of parallel fringes seen beyond CD which is characteristic of the contrast from single stacking faults. On the other side of CD the contrast over ACD is dotted and of exactly the same nature as that seen within the tetrahedron K. The Burgers vectors of the dislocations of the defect in the configuration of fig. 10 (*c*) are shown in fig. 10 (*e*). It must be emphasized that the loss of contrast in plane DCB cannot be due to the mechanism invoked in § 3 as the (200) reflections operating in this orientation are affected by all stacking faults (Paterson 1952, Whelan and Hirsch 1957 a, b).

In figs. 12 and 13 sets of loops can be observed which correspond to equilateral triangles on (111) planes with edges parallel to [110] projections. It is thought that these are probably due to the collapse of tetrahedra to form Frank sessiles, but it is not clear at present why these are stable. Some of these loops show definite imperfections along their

† Figures 10 (*d*) and (*e*) have been rotated by 90° in an anti-clockwise direction relative to figs. 10 (*a*), (*b*) and (c).

edges. Occasionally glissile prismatic loops can be observed, as for example at L in figs. 12, 13.

Figure 14 shows an example where a glissile dislocation (probably a Shockley partial) emerges from a defect, R.

5.4. *Interaction of Dislocations with the Defect*

The defects are expected to interact with dislocations and to present obstacles to their movement. Figures 9 (*a*, *b*) show a sequence in which a screw dislocation interacts with a tetrahedron. Figure 9 (*a*) shows the dislocation M parallel to a [110] direction, and from its subsequent cross-slip trace (fig. 9 (*b*)) it may be deduced that the dislocation must be a screw dislocation. Comparing figs. 9 (*a*) and 9 (*b*) it appears that the tetrahedron collapses on to a (111) plane as a result of the passage of the screw. The mechanism by which this process occurs is not understood at present; however, the sequence suggests that the screw cross-slips around the obstacle in a way similar to that observed for age-hardened alloys, (Thomas *et al.* 1957), indicating that there is a repulsion between the screw and the defect. The sequence is also consistent with the view that the mutual repulsion causes the defect to collapse on a plane furthest away from the screw. There is, however, also the possibility that the screw actually passes through the tetrahedron and causes it to be truncated, as explained below.

Figures 10 (*a*, *b*, *c*) show another example of an interaction between a screw dislocation and a defect. In fig. 10 (*a*) a truncated tetrahedron can be seen clearly at N; in fig. 10 (*b*) a slip trace has appeared which ends at N, and appears to be terminated by a dislocation. In fig. 10 (*c*) the dislocation and slip trace have disappeared and a very small tetrahedron now replaces the previously larger truncated defect. The slip trace in fig. 10 (*b*) is approximately parallel to a (110) plane and presumably corresponds to the average path of a screw dislocation undergoing cross-slip. The situation is therefore very similar to the case discussed above. Suppose now that the approaching screw, T, lies parallel to AC (with Burgers vector **CA**), and that it moves in a plane parallel to ABC, on the side of ABC on which D lies, fig. 10 (*f*). (To simplify the discussion the tetrahedron is considered to be complete.) To understand this interaction completely, it should be discussed in terms of the interaction of the partials of the screw with the tetrahedron; the details of this analysis have still to be worked out. However, the result of the interaction can be arrived at as follows. Suppose the tetrahedron collapses back on to the plane ABD to form a Frank sessile Cγ, the dislocation now advances to intersect the triangle ABD along EF (fig. 10 (*g*)), and then splits into two dislocations according to the reaction

$$\mathbf{CA} = \mathbf{C\gamma} + \mathbf{\gamma A}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The dislocation $\mathbf{\gamma A}$ now sweeps across the plane ABEF, removing the fault and combining with the Frank sessile FABE to form a dislocation

with Burgers vector **CA** according to the reverse of the reaction (6). The resulting dislocation GFABEH (fig. 10 (*h*)) with Burgers vector **CA** has all its segments lying in (111) planes except for EB, which is an edge dislocation in a (100) plane. This part of the dislocation will not be able to glide easily, particularly at low temperatures, and will therefore impede the movement of the dislocation and cause hardening. Under the conditions of the high stresses in the foils the dislocation is apparently able to glide away, as is evident from fig. 10 (*c*). The reaction is equivalent to a combination of the dislocation GFEH with that part of the dislocation loop corresponding to those vacancies on the compression side of the dislocation. The remaining part of the defect left behind, i.e. the triangle EDF (fig. 10 (*h*)) dissociates again to form a corresponding smaller tetrahedron seen at O in fig. 10 (*c*).

For a screw dislocation of opposite sign a similar reaction is expected, except that now the dislocation **γA** will sweep across the triangle FDE, resulting in the formation of a whole dislocation GFDEH and a sessile ring AFEB. The latter may dissociate again to form a truncated tetrahedron, and it is possible that this reaction corresponds to that observed in figs. 9 (*a*, *b*).

It is also of interest to note here that in fig. 9 (*b*) the defect P appears particularly distorted by the strain field of the nearby dislocation Q. The slighter distortions of some of the other defects are probably due to high local stresses in this area.

The above examples suggest that the defects cause hardening by impeding the motion of the dislocations. However, more experimental and theoretical studies are required to establish the details. Hardening in quenched Au has been reported by Kauffmann and Meshii (1957) and interpreted as due to sessile rings by Kimura *et al.* (1958 a, b, c, see also Maddin *et al.* 1958).

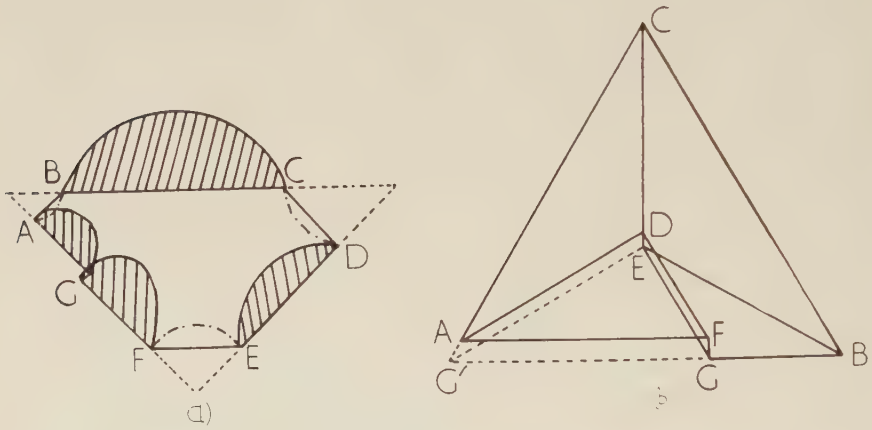
§ 6. Discussion

The results show clearly that the defects occurring in quenched Au are tetrahedra of stacking faults. The facts that these defects occur only in quenched samples, that they can be derived from collapsed triangular discs of vacancies, and that there is good agreement between the values of vacancy concentration corresponding to the defects and those deduced from resistivity measurements, leave little doubt that the tetrahedra are formed as a result of the condensation of vacancies.

It was shown in § 4 that triangles of Frank sessiles on (111) planes, with their edges parallel to [110] directions would automatically dissociate into tetrahedra to lower the energy. Although the sides of the Frank sessile loops are expected to lie along [110] directions since the dislocations can then dissociate, thereby lowering the energy, when the loops are first formed under conditions of very high supersaturation by the collapse of vacancy discs, they are probably nearly circular. Under these conditions only very short lengths of a loop lie along close-packed

directions, i.e. a section of the loop may be considered as a heavily jogged dislocation line lying along a close-packed direction. Any dissociation will therefore be suppressed as the jog energy will be much greater when the dislocation is dissociated. As the loop grows and as the concentration of vacancies decreases there will be a tendency for the sides of the loops to lie along [110] directions so that dissociation can take place. In the next stage of the process the loop will therefore be more or less hexagonal, the sides of the hexagons being parallel to [110] directions but containing a number of jogs. Such an intermediate state is shown schematically in fig. 15 (*a*). The partial dislocations above the plane of the loop attract each other, and so do

Fig. 15



(*a*) The intermediate stage of growth between the circular Frank sessile and the tetrahedron of stacking faults. Notice the jog G in the irregular hexagon ABCDEF. The partials bounding shaded faults bow out above the plane ABCDEFG and those represented by the dash-dotted lines bow out below this plane.

(*b*) A completed tetrahedron showing the extended step EDFG in the fault caused by the jog GF in the stair-rod AB. The final configuration BG′EC after climb of the jog has taken place is outlined by the dashed lines G′G, G′A, G′E, and by the full lines BG, BC, BD, CE and CA.

those below the plane. The hexagons are not expected to be regular; if for example AB is relatively short the partials from BC and FA will attract each other particularly strongly, and the resultant stress field drives the vacancies to AB till an acute angled corner is formed. A similar process at CD and FE finally leads to the formation of a triangle and of the complete tetrahedron. If the edges of the triangle contain jogs, the tetrahedron will contain extended jogs which run across its faces; an example is shown in fig. 15 (*b*). Vacancies will be attracted to such jogs FG on the stair-rod AB; the extended step EDFG in the fault consists of

two Shockley partials of opposite sign which can glide on the planes ADF and EGB respectively as the jog advances. When the jog reaches the corner A the Shockley partial previously along DF reacts with the stair-rod AD to form a Shockley partial on plane CDA which glides and reacts with the partials EG′ and DE to form the stair-rods along these directions. In this way the face of the tetrahedron moves outward, and the final stages of the growth of the tetrahedron will take place by this process.

The energy of such extended jogs will be high and there will therefore be a strong tendency for all jogs to disappear eventually, thus leaving perfect tetrahedra. Once a perfect tetrahedron is formed it can only grow by the formation of new jogs. Under conditions of high supersaturation or at high temperatures jogs might be nucleated at the corners thus leading either to growth or dispersal of the tetrahedra.

For a metal of high stacking fault energy a complete prismatic dislocation loop will be formed (Read 1953, Kuhlmann-Wilsdorf 1958, Hirsch *et al.* 1958). The sides of the loops will again be expected to lie along [110] directions, since some dissociation can take place for dislocations lying along these directions. The dislocations along four sides of the hexagon lie in (111) planes and these can dissociate into Shockley partials; along the remaining two sides the dislocations are lying in (100) planes, and these can dissociate into a low energy stair-rod plus two Shockley partials on the two (111) planes intersecting in the [110] direction. Since the dissociation is limited, there is little attraction between the partials from the different segments of the hexagon and the final configuration is expected to be hexagonal rather than triangular. Actually, in Al many of the loops appear to have hexagonal or distorted hexagonal shapes, although some parallelogram shaped loops have also been observed (Hirsch *et al.* 1958).

It seems that the model proposed by Kimura *et al.* (1958 a, b, c) for the annealing out of vacancies in metals of low stacking fault energy is basically correct in that stacking faults are produced. The sessile rings assumed in their analysis should, however, be replaced by the tetrahedra, and it follows at once that for a given number of vacancies the area of stacking fault produced is four times as large as that assumed by these authors. In the present experiments the average area of stacking fault per defect is about $2{\cdot}1 \times 10^5$ Å^2, equivalent to a circular area of diameter 520 Å; this is considerably smaller than the estimate of ~ 7000 Å made by Kimura *et al.*, which however is an upper limit. The number of defects observed is about 5×10^{14} cc^{-1}, which is considerably larger than their estimate of $1{\cdot}4 \times 10^{11}$ cc^{-1}. It is not clear yet whether these magnitudes are consistent with those predicted by their theory.

With regard to the problem of quench hardening in metals of low stacking fault energy (Kauffman and Meshii 1957, Kimura *et al.* 1958 a, b, c, Maddin *et al.* 1958) it is clear from § 5.4 that the defects present obstacles to slip by other dislocations. The presence of large numbers of jogs on these dislocations may also lead to hardening. On the other hand, if no

other dislocations are present, glissile dislocations could be nucleated at the defects under the influence of high stresses. This again leads to hardening, and probably also to a yield point. More experiments are required to determine which of these mechanisms is the more important. Experiments on the variation of quenching temperature and quenching rate should also give information about the type and arrangement of defects and of dislocations appropriate to a given treatment.

Finally, it should be noted that defects of any kind (including small cavities) below about 50–100 Å in size are difficult to detect and their presence cannot be ruled out.

Acknowledgments

Our thanks are due to Professor N. F. Mott and Dr. W. H. Taylor for their continued interest and encouragement, to Professor Maddin for suggesting the experiments on Au, to Professor Koehler whose stimulating course of lectures aroused our interest in the quenching problem, to Dr. Kimura and to our colleagues for helpful discussions, and to C. K. Jackson for unfailing attention to the performance of the microscope.

One of us (J.S.) is indebted to A.E.R.E. (Harwell) for a maintenance grant.

APPENDIX

Maximum Size of Tetrahedron

If the side of the tetrahedron, l, is large the energy of the total area of stacking faults becomes more important than the line energy of the dislocations; under these conditions the Frank sessile loop corresponds to the configuration of lower energy. The maximum size of the tetrahedron can then be found as follows:

The energy of a triangle of Frank partials is

$$E_{\mathrm{L}} = \frac{Ga^2 l}{4\pi(1-\nu)} \log \frac{l}{r_0} + \frac{\sqrt{3}}{4} l^2 \gamma$$

where G is the shear modulus, a is the unit cell size, r_0 is the core radius of the dislocation, ν is Poisson's ratio, and γ is the stacking fault energy.

The energy of the tetrahedron is

$$E_{\mathrm{T}} = \frac{Ga^2 l}{12\pi(1-\nu)} \log \frac{l}{r_0} + \sqrt{3} l^2 \gamma .$$

The maximum size l_0 of a tetrahedron is given by the condition $E_{\mathrm{L}} = E_{\mathrm{T}}$, i.e.

$$\frac{l_0}{\log l_0/r_0} = \frac{2Ga^2}{9\sqrt{3}\pi(1-\nu)\gamma} ;$$

taking $G = 2{\cdot}7 \times 10^{11}$ dyn/cm^2, $\gamma = 33$ ergs/cm^2, $\nu = 0{\cdot}3$, and $r_0 \sim \frac{1}{2}b$, where b is the Burgers vector of a $\frac{1}{2}[110]$ dislocation, we find $l_0/b \sim 150$, i.e. $l_0 \sim 430$ Å. This calculation is only approximate and it should be noted that the energy of the Frank sessile loop will be lower since the Frank partials will in any case dissociate to form ribbons of fault on the (111) planes on which for l smaller than l_0 the complete tetrahedron is formed.

References

Amelinckx, S., 1957, *Report of Lake Placid Conference on Dislocations and Mechanical Properties of Crystals* (New York: Wiley), p. 3.

Barnes, R. S., 1954, *Acta Metall.*, **2,** 380.

Bauerle, J. E., and Koehler, J. S., 1957, *Phys. Rev.*, **107,** 1493.

Fisher, J. C., 1957, *Report of Lake Placid Conference on Dislocations and Mechanical Properties of Crystals* (New York: Wiley,) p. 513.

Frank, F. C., 1950, *Carnegie Institute of Technology (Pittsburgh) Symposium on the Plastic Deformation of Crystalline Solids, O.N.R.*, p. 150; 1956, *Deformation and Flow of Solids* (Berlin : Springer), p. 73.

Friedel, J., 1955, *Phil. Mag.*, **46,** 1169.

Gilman, J. J., and Johnston, W. G., 1957, *Report on Lake Placid Conference on Dislocations and Mechanical Properties of Crystals* (New York: Wiley), p. 116.

Hirsch, P. B., Horne, R. W., and Whelan, M. J., 1956, *Phil. Mag.*, **1,** 677.

Hirsch, P. B., Silcox, J., Smallman, R. E., and Westmacott, K. H., 1958, *Phil. Mag.*, **3,** 897.

Jones, D. A., and Mitchell, J. W., 1958, *Phil. Mag.*, **3,** 1.

Kauffman, J. W., and Meshii, M., 1957, *Bull. Amer. phys. Soc.*, **2,** 145.

Kimura, H., Maddin, R., and Kuhlmann-Wilsdorf, D., 1958 a, *Bull. Amer. phys. Soc.*, **3,** 122; 1958 b, *Acta Metall.* (in the press); 1958 c, *J. Metals*, **10,** 93.

Koehler, J. S., Seitz, F., and Bauerle, J. E., 1957, *Phys. Rev.*, **107,** 1499.

Kuhlmann-Wilsdorf, D., 1958, *Phil. Mag.*, **3,** 125.

Maddin, R., Kimura, H., and Kuhlmann-Wilsdorf, D., 1958, *Bull. Amer. phys. Soc.*, **3,** 125.

Paterson, M. S., 1952, *J. appl. Phys.*, **23,** 805.

Read, W. T., 1953, *Dislocations in Crystals* (New York: McGraw-Hill).

Seeger, A., 1957, *Report of Lake Placid Conference on Dislocations and Mechanical Properties of Crystals* (New York: Wiley), p. 243.

Seitz, F., 1950, *Phys. Rev.*, **79,** 890.

Suzuki, H., and Barrett, C. S., 1958, *Acta Metall.*, **6,** 156.

Teghtsoonian, E., and Chalmers, B., 1951, *Canad. J. Phys.*, **29,** 370; 1952, *Ibid.*, **30,** 388.

Thomas, G., Nutting, J., and Hirsch, P. B., 1957, *J. Inst. Metals*, **86,** 7.

Thompson, N., 1953, *Proc. phys. Soc. Lond.* B, **66,** 481; 1955, *Report of Conference on Defects in Crystalline Solids* (London: Physical Society), p. 153.

Thornton, P. R., and Hirsch, P. B., 1958, *Phil. Mag.*, **3,** 738.

Whelan, M. J., 1958 a, *Ph.D. Dissertation*, University of Cambridge; 1958 b, *Proc. roy. Soc.* (in the press).

Whelan, M. J., and Hirsch, P. B., 1957 a, *Phil. Mag.*, **2,** 1121; 1957 b, *Ibid.*, **2,** 1303.

Whelan, M. J., Hirsch, P. B., Horne, R. W., and Bollmann, W., 1957, *Proc. roy. Soc.* A, **240,** 524.

Cosmic Ray Cut-off Rigidities and the Earth's Magnetic Field†

By J. J. Quenby and W. R. Webber
Imperial College of Science and Technology, London

[Received September 15, 1958]

Abstract

Approximate values of the vertical cut-off rigidities for cosmic ray particles in the earth's magnetic field have been deduced taking into account both the dipole and the non-dipole parts of the internal field. The accuracy of these calculated values is discussed and it is shown that they fit the experimental data rather well.

§ 1. Introduction

Accumulated evidence during the past two or three years has shown that there are serious discrepancies between the observed distribution of cosmic ray intensity over the earth and that which would be expected on the basis of the usual representation of the earth's magnetic field as that of a dipole. These discrepancies are evident from neutron latitude surveys carried out both at sea level (Rose *et al.* 1956, Kodama and Miyazaki 1957, Rothwell and Quenby 1957, Skorka 1958) and at aeroplane altitudes (Simpson 1957), and were particularly clearly shown during the solar flare of February 23, 1956, when the intensity of the additional isotropic radiation was not distributed over the earth's surface in a way which could be reconciled with the usual eccentric dipole representation of the field (Pfotzer 1957). Further evidence of these difficulties is to be found in the disagreement between the measured and predicted values of cut-off rigidity for α-particles (McDonald 1957, Waddington 1956). Two possible explanations of these discrepancies have been suggested. On the one hand, that they arise from the distortion of the geomagnetic field at large distances by conducting interplanetary gas (Simpson *et al.* 1956), and on the other, that they are due to the regional anomalies in the field (Rothwell and Quenby 1957).

It was pointed out long ago (see, for example, Johnson 1938) that the cosmic ray intensity at sea level was closely related to the local value of the earth's field and this relationship has been clearly shown for the South African magnetic anomaly by Rothwell and Quenby (1957). It is evident from these results that the simple dipole representation of the earth's field is inadequate for a detailed discussion of cosmic ray phenomena and we must, therefore, take account of the non-dipole part of the field

† Communicated by the Authors.

which is of internal origin before attempting to discuss the possible effects of unknown fields of external origin. In principle, this could be done by computing large numbers of trajectories of cosmic ray particles in the earth's field, but this is a formidable task, especially if it is to be done for a large number of places. It seems worth while, therefore, to try to find an approximate treatment which is sufficiently good for most purposes and yet does not involve a disproportionate amount of computation. An empirical expression which leads to more accurate values of the cut-off rigidities than those for a centred dipole has already been given by Rothwell (1958).

In the present paper we consider the general properties of the trajectories of charged particles in a dipole field and deduce approximate corrections to the cut-off rigidities which must be applied in order to take into account effects of the non-dipole parts of the internal field. A series of tables and maps has been prepared showing the values of the vertical cut-off rigidity so deduced for every 10° of latitude and 15° of longitude. The accuracy of the calculated values is discussed and it is shown that they fit the experimental data rather well.

§ 2. Stormer Treatment of the Motion of Charged Particles in a Dipole Field

Since we intend to appeal qualitatively to certain aspects of the motion of particles in a dipole field, it is worth while recalling some of the basic features of Störmer theory at this stage. In discussing the equation of motion of a particle in such a field, it is convenient to work in spherical polar coordinates with the centre of the earth as the origin and with the dipole directed along the negative z direction. r is the distance from the origin and λ and w are respectively the geomagnetic latitude and west geomagnetic longitude. The w direction resolute of the Lagrangian equation of motion for a charged particle moving in a static magnetic field is

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{w}} = \frac{\partial L}{\partial w} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where L, the Lagrangian of the particle, is given by

$$L = -mc^2\sqrt{\left(1-\frac{v^2}{c^2}\right)} + \frac{e}{c}\mathbf{A}\,.\,\mathbf{v} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where m is the rest mass, $\mathbf{v}$ the velocity and e the charge of the particle and $\mathbf{A}$ is the vector potential describing the magnetic field. In a dipole field, the vector potential may be written as

$$\mathbf{A} = \mathbf{i}_w \frac{M\cos\lambda}{r^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

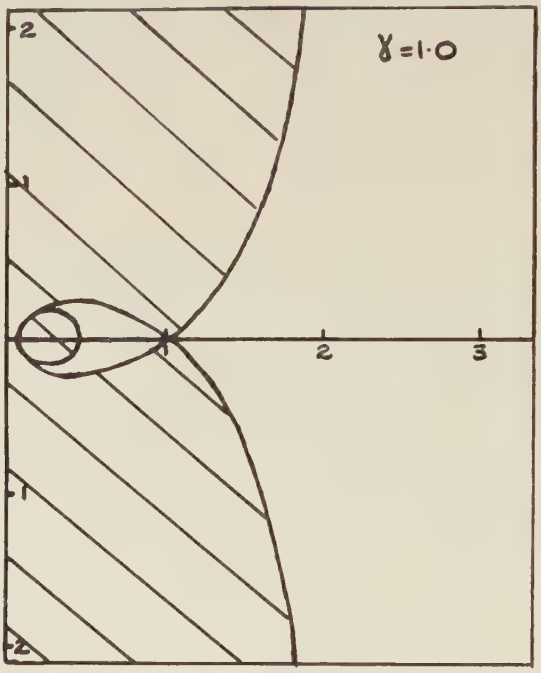
where $\mathbf{i}_w$ is a unit vector in the w direction, i.e. pointing west, and M is the dipole moment. Since $\mathbf{A}$ is not a function of w, eqn. (1) may be integrated (see, for example, Fermi 1949), leading to the well-known

relation obtained by Störmer for the motion of a charged particle in the meridian plane, defined here as the plane through the moving particle and the z axis.

$$b = r\cos\lambda\sin\theta + \frac{Ar\cos\lambda}{P}, \quad \ldots\ldots \quad (4)$$

In eqn. (4) b is a constant with dimensions of length, θ is the angle between the velocity vector of the particle and the meridian plane containing the particle and is positive if the particle crosses the meridian plane from east to west. $P = pc/e$ where p is the particle momentum. If we substitute

Fig. 1



Allowed and forbidden regions in the meridian plane for a dipole field ($\gamma = 1$).

for A and use the Störmer unit of length, $C = \sqrt{(M/p)}$ defined by putting $r = \sqrt{(M/p)}R$ in eqn. (4), we get the dimensionless equation

$$2\gamma = R\cos\lambda\sin\theta + \frac{\cos^2\lambda}{R} \quad \ldots\ldots \quad (5)$$

where γ is a constant. For a given value of γ, areas in the meridian plane where $|\sin\theta| > 1$ are forbidden to the particle, the boundary line between the allowed and forbidden regions being obtained by putting $\sin\theta = 1$ in eqn. (5). In general there are two types of allowed region ; the outer one extending to infinity and the inner one near the origin which may or may not be connected to the outer region, depending on the value of γ. Figure 1 shows the allowed and forbidden regions for $\gamma = 1$. Here the

inner allowed region is separated from the outer by the jaws of the forbidden region which just meet in the equatorial plane. For $\gamma < 1$ the jaws are open and the two allowed regions connect, while for $\gamma > 1$ the inner and outer allowed regions are completely separate. In the diagram the radius of the earth, R_e is measured in Störmer units and therefore depends on the particle momentum. If $R_e < 1$ the surface of the earth lies in the inner allowed region and particles of corresponding momentum can only reach the earth provided $\gamma < 1$. The limiting condition for the minimum rigidity a particle must have to arrive at the earth is given by $\gamma = 1$. Substituting this in eqn. (5) leads to the expression for the cut-off rigidity in Bev/c. At an angle θ this is

$$P = \frac{M}{r_e^2}\left[\frac{1-\sqrt{(1-\sin\theta\cos^3\lambda)}}{\sin\theta\cos\lambda}\right]^2. \quad . \quad . \quad . \quad . \quad (6)$$

For vertical arrival, $\sin\theta = 0$ and this expression becomes

$$P = \frac{M}{4r_e^2}\cos^4\lambda. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

In considering the cut-off rigidities in the earth's field we shall assume that the above theory still applies approximately and deduce corrections for the effect of the non-dipole parts of the real field which are neglected in the Störmer treatment.

§ 3. The Earth's Magnetic Field

The magnetic potential of the earth may be represented to any required degree of accuracy by the use of spherical harmonics. Thus we may write the geomagnetic potential as

$$V = \sum_{n=1}^{\infty} V_n,$$

$$V_n = r_e\left(\frac{r_e}{r}\right)^{n+1} T_n \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where r is the distance from the centre of the earth, r_e is the radius of the earth and

$$T_n = \sum_{m=0}^{n} (g_n^m \cos mw + h_n^m \sin mw) P_n^m(\theta) \quad . \quad . \quad . \quad (9)$$

where g_n^m and h_n^m are the Gauss coefficients, w is the geographic longitude, θ is the geographic colatitude and $P_n^m(\theta)$ are the normalized spherical harmonics introduced by Schmidt (1934). The squared value of T_n averaged over the complete spherical surface is given by

$$|T_n|^2 = \sum_{m=0}^{n} [(g_n^m)^2 + (h_n^m)^2]/2n+1. \quad . \quad . \quad . \quad . \quad (10)$$

In practice the coefficients g_n^m, h_n^m are chosen so as to give a best fit between grad V of eqn. (8) and the measured field. These coefficients for terms up to $n = 6$ have been determined from the 1955 magnetic data by Finch and Leaton (1957).

The potential terms with $n=1$ in eqn. (8) correspond to a dipole which is situated at the geographic centre of the earth and tilted with respect to the axis of rotation. In this paper we shall start from the cut-off rigidities which apply in the dipole field alone and then make corrections to take into account the effect of the potential terms with $n>1$. It is therefore of interest to examine the relative importance of the higher order terms with respect to the dipole terms.

Using the values of g_n^m, h_n^m given by Finch and Leaton for 1955, the root mean square values of T_n have been calculated. Values of $|V_n|$ expressed as percentages of $|V_1|$ are given as a function of distance from the centre of the earth in table 1.

Table 1. Relative Importance of Various Spherical Harmonic Terms

	% of $\|V_1\|$					
Distance	$\|V_2\|$	$\|V_3\|$	$\|V_4\|$	$\|V_5\|$	$\|V_6\|$	$\sum_{n=1}^{6} \|V_n\|$
$1{\cdot}0\ r_e$	10·4	5·9	2·8	0·9	0·4	20·4
$1{\cdot}2\ r_e$	8·7	4·1	1·6	0·4	0·2	15·0
$1{\cdot}5\ r_e$	6·8	2·6	0·8	0·2	0·1	10·5
$2{\cdot}0\ r_e$	5·2	1·5	0·3	0·1	$<0{\cdot}1$	7·0
$3{\cdot}0\ r_e$	3·5	0·7	0·1	$<0{\cdot}1$	$<0{\cdot}1$	4·2

The higher order terms considered here amount to only 20% of the dipole contribution to the geomagnetic potential at the surface and to less than this at points outside the surface.

§ 4. Derivation of Approximate Expressions for Cut-off Rigidities in the Earth's Magnetic Field

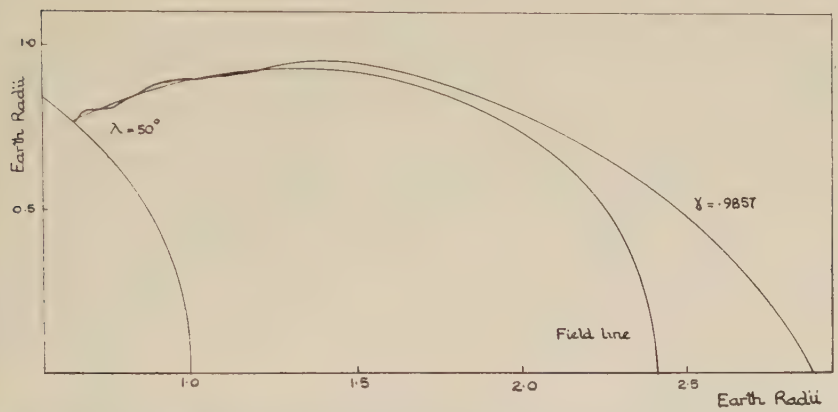
In the discussion which follows we suppose that to a first approximation cosmic rays behave in the earth's field as they would in that of a dipole and we then consider the perturbing effects on the trajectories of the non-dipole parts of the field. A similar approach has been used by Jory (1956) in discussing the effect of the quadrupole terms on the cut-off rigidities near the geomagnetic equator. In the present paper we shall consider the effects of all the significant non-dipole terms at both high and low latitudes.

We start by considering the motion of particles arriving at high latitudes. In this case the radius of the earth measured in Störmer units for vertical cut-off particles is less than $\frac{1}{2}$ and the point of arrival (r_e, λ) therefore appears in fig. 1 inside the horn-shaped part of the inner allowed region.

We may conveniently divide the trajectory of a particle moving in the meridian plane just above the cut-off rigidity into three sections. The first is the passage in the equatorial plane region through the narrow gap

in the jaws of the forbidden areas which exists for γ just less than unity. The second is the complicated series of loops that the particle performs in the main part of the inner allowed region just inside the jaws. The third is the motion up the narrowing horn-shaped region to the point of observation. In this third stage of the motion the particle has a high degree of contact with the line of force through the point of observation. This important property of the trajectories has been pointed out by Störmer (1955) who has given many examples. Figure 2 shows the beginning of

Fig. 2



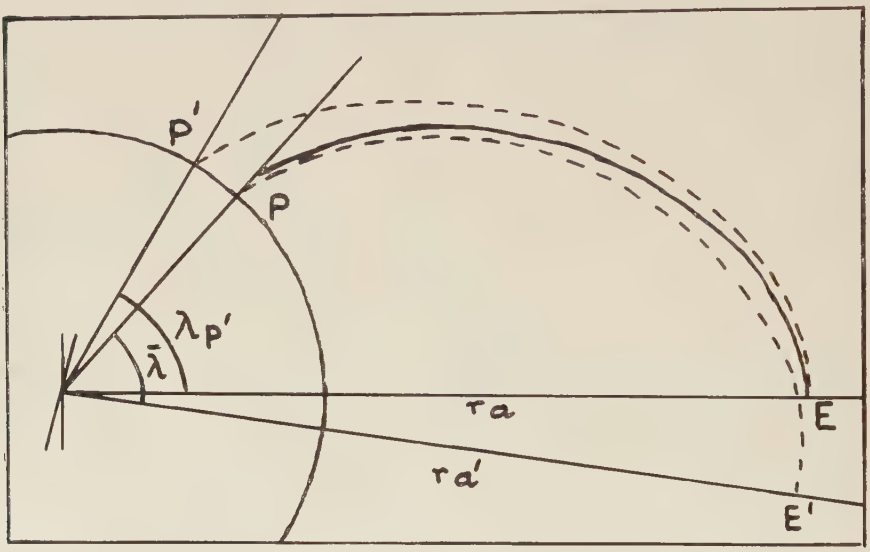
Trajectory of vertically arriving particle just above cut-off rigidity at $\lambda = 50°$. Note close contact with the line of force in the vicinity of the earth.

the trajectory of a negative particle leaving the earth at $\lambda = 50°$ with a rigidity just above cut-off. In the vicinity of the earth the high degree of contact with the line of force starting from the same point is clearly seen.

At latitudes lower than about 30° the surface of the earth in fig. 1 corresponding to the vertical cut-off particle has moved out of the horn-shaped part of the inner allowed region and the particles near cut-off now no longer follow closely the line of force through the point of observation. For the present then we limit our discussion to latitudes greater than 40°, where there is close contact between the field lines and the particle trajectories and consider first of all the entry of particles through the equatorial jaws. The effect of the non-dipole parts of the field will be to change the critical value of γ at which the jaws just close in the equatorial plane. In order to allow for the distortion of the jaws and to apply a correction for this, we should need to know the point at which the trajectory passes through the jaws. This can only be found by integrating the equations of motion. If we restrict our treatment to latitudes greater than 40°, however, this difficulty can be avoided because the correction due to the effect of the non-dipole field on the jaws is sufficiently small to be neglected.

This can be shown as follows. In a dipole field the jaws lie at a distance $2R_e/\cos^2\lambda$ from the dipole. This distance varies from $2R_e$ for particles arriving at the equator ($\lambda = 0$) to infinity for the particles arriving at the pole ($\lambda = 90°$). At latitude 40° the jaws lie at a distance of 3·4 r_e from the centre of the earth and at this distance the contribution of the non-dipole terms introduces an error which probably amounts to less than 3% in the cut-off rigidity (see table 1). Since the contribution of the non-dipole terms decreases with increasing latitude as the jaws move further out, the errors introduced by neglecting the effect of these terms soon become negligible.

Fig. 3



Schematic picture of the line of force argument at high latitudes. Real line of force PE (solid line) with the dipole line of best fit PE′ and the dipole line asymptotic to the real line, P′E. (dotted lines).

That part of the trajectory which lies between the jaws and the final approach along a line of force is in general very complicated and varies widely for small changes in momentum. Conversely this means that perturbations in this region, even though they produce large changes in the trajectory, have little effect on the cut-off rigidity at any point. It is in fact the last part of the trajectory which is crucial since it is this which determines where particles of a given momentum actually arrive at the earth's surface. We will therefore assume that prior to the entry of the particles into the last stage of their trajectory, they are behaving as they would in a pure dipole field.

Let us now examine the effect of the non-dipole part of the field on the last stages of the trajectory. Figure 3 shows a line of force connecting

the point P on the earth to a point E in the equatorial plane. The dotted line represents a line of force starting from E as it would be if we had a purely dipole field. We now imagine a particle (of near cut-off rigidity) travelling towards the earth's surface along the line of force from E. If the field were entirely dipole in character it should arrive at P′ but because of the deviation of the true line from the dipole line it arrives instead at the point P. Consequently we assume that the cut-off rigidity at P will be that which would be appropriate to P′ if the field were a perfect dipole. That is to say the cut-off rigidity at P would be given by

$$P = \frac{M}{4r_e^2}\cos^4\lambda_p' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where λ_p' is the geomagnetic latitude of P′.

It follows then that if we can find the latitude λ_p' of P′ we can find the cut-off at P by using the Störmer expression given above. In other words, we wish to identify the dipole field line which is asymptotic to the real line in the equatorial plane. Instead of computing the field outside the earth from the spherical harmonic expansion of the surface field, a laborious process which is probably not justified in view of the schematic nature of our treatment, we adopt the following approximate method based on the readily available charts of the surface field parameters. These charts in effect give the sum of the first six spherical harmonic terms and represent a smoothed version of the actual surface field.

We replace the real line of force by another line of force PE′ generated by a hypothetical dipole. This dipole has an angle of tilt relative to the geomagnetic dipole adjusted so as to give a best fit to the horizontal and vertical components of the real field along a radius through P. We then use the equation of this hypothetical dipole line to trace the real line into the equatorial plane and hence identify the required dipole line P′E. If $\bar{H}$ and $\bar{V}$ are the horizontal and vertical components of the field due to our hypothetical dipole, the conditions for best fit to the real fields are

$$\int_{r_e}^{\infty}\frac{\bar{H}r_e^3}{r^3}dr = \int_{r_e}^{\infty}\frac{H_c r_e^3}{r^3}dr + \int_{r_e}^{\infty}\frac{\Delta H r_e^4}{r^4}dr \quad . \quad . \quad . \quad (12a)$$

and

$$\int_{r_e}^{\infty}\frac{\bar{V}r_e^3}{r^3}dr = \int_{r_e}^{\infty}\frac{V_c r_e^3}{r^3}dr + \int_{r_e}^{\infty}\frac{\Delta V r_e}{r^4}dr \quad . \quad . \quad . \quad (12b)$$

where H_c and V_c are the surface field components of the geomagnetic dipole field at P and $\Delta H = H_A - H_c$, $\Delta V = V_A - V_c$ represent the non-dipole parts of the real field at P, taken to be entirely quadrupole in this instance. On integration we find

$$\bar{H} = H_c + \tfrac{2}{3}\Delta H, \quad \bar{V} = V_c + \tfrac{2}{3}\Delta V. \quad . \quad . \quad . \quad (13a, b)$$

In other words the non-dipole field when assumed to be all quadrupole at the point P receives a weight of $\frac{2}{3}$ that of the regular dipole field.

Similar integrations assuming non-dipole fields falling off as r^{-5}, r^{-6}, r^{-7} and r^{-8} yield factors of $\frac{1}{2}$, $\frac{2}{5}$, $\frac{1}{3}$ and $\frac{2}{7}$ respectively instead of $\frac{2}{3}$. If these

factors are now weighted according to the average world wide contribution of the various order terms to the field a mean weighted factor of 0·52 is obtained.

Strictly speaking we should use a weighted average obtained by taking into account all the terms in the infinite series of spherical harmonics. We have, however, used only terms up to the 6th order for two reasons. In the first place the effect of higher order terms is quite insignificant and secondly there are readily available charts of the surface field constructed by using harmonics up to the 6th order.

The equation of our hypothetical dipole line of force PE′ can now be written as

$$r' = r_a' \cos^2 \lambda' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

where primed quantities are measured with respect to the tilted dipole, and where $r_a' = r_e/\cos^2 \bar{\lambda}$, r_a' being the distance between the hypothetical dipole and the point of intersection of the line of force with its equatorial plane. $\bar{\lambda}$ is the latitude of P with respect to the tilted dipole.

$$\bar{\lambda} = \tan^{-1}\left(\frac{\bar{V}}{2\bar{H}}\right) = \tan^{-1}\left(\frac{V_c + 0{\cdot}52\Delta V}{2(H_c + 0{\cdot}52\Delta H)}\right). \quad . \quad . \quad . \quad (15)$$

Similarly the equation of the geomagnetic field line passing through P′ can be written

$$r = r_a \cos^2 \lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where

$$r_a = \frac{r_e}{\cos^2 \lambda_{p'}}$$

where r_a and $\lambda_{p'}$ are corresponding quantities for the geomagnetic dipole. Provided that the tilt of the hypothetical dipole relative to the geomagnetic dipole is small we can write $r_a = r_{a'}$. We then have $\lambda_{p'} = \bar{\lambda}$ which now identifies the dipole line of force through P′ as that corresponding to a dipole latitude $\bar{\lambda}$. Consequently the cut-off rigidity at P is given by

$$P_M = \frac{M}{4r_e^2} \cos^4 \bar{\lambda} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

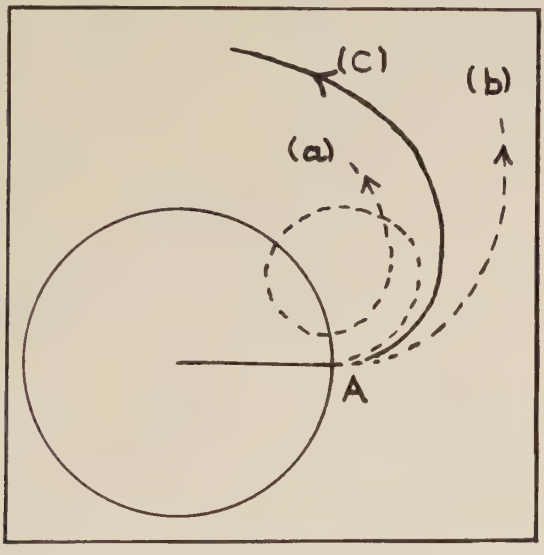
using the usual Störmer expression at a latitude $\bar{\lambda}$.

The discussion given above is valid at latitudes greater than 40°. At low latitudes however the particles of near cut-off rigidity no longer follow lines of force into the horns of the inner allowed region and our treatment therefore breaks down. In addition the equatorial jaws are much closer to the surface of the earth and distortions of these jaws due to non-dipole terms becomes increasingly important.

In discussing the trajectories at low latitudes it is simplest to begin by considering those in the equatorial plane. In this plane the trajectories being two dimensional, are relatively simple and in discussing them it is helpful to think in terms of a negative particle emitted from the earth rather than a positive particle coming in from infinity. Such a negative

particle traces out the same path as a positive particle of the same magnetic rigidity coming in. The cut-off rigidity is then that which the negative particle must have to escape to infinity. Consider such a particle emitted vertically from the point A which lies in the equatorial plane of a dipole whose axis is perpendicular to the paper. (see fig. 4) If the particle has a rigidity less than that corresponding to the critical value it will travel in a trajectory of type (*a*) and will not escape. On the other hand, if the particle has a rigidity greater than the critical value it will escape along a trajectory (*b*). The limiting case is the circular periodic orbit (*c*) which has a radius $2r_e$ where r_e is the radius of the earth. A decrease in the field strength in the region of A due to non-dipole terms in the real field potential may change a trajectory of type (*a*) into one of type (*b*); that is to say it results in a decrease in the cut-off rigidity whilst an increase in field strength has the opposite effect.

Fig. 4



Types of orbits in the equatorial plane of a dipole.

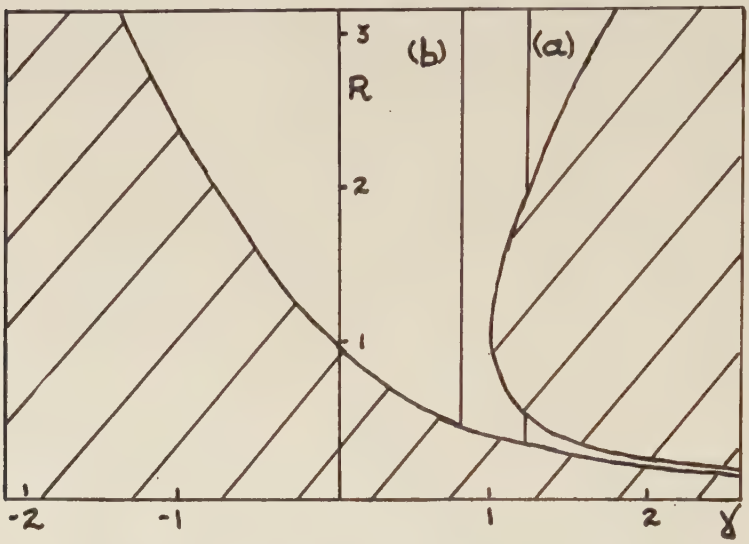
We can now make a quantitative estimate of the effect of the non-dipole terms of the field potential on the cut-off rigidities in the equatorial plane. In order to do so, we assume that to a first approximation the force deflecting the particle out of the equatorial plane can be neglected. In addition we shall neglect the dependence of the quadrupole and higher order terms on the azimuth angle w. This second assumption seems not unreasonable in the light of the qualitative discussion above where we see that it is primarily the non-dipole field in the immediate vicinity of A which determines whether or not the trajectory goes to infinity. Consider the particular case of a particle moving entirely in the equatorial plane

under the combined influence of a dipole and quadrupole field. The vector potential in the equatorial plane can then be written

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 = \mathbf{i}_w \frac{M}{r^2} + \mathbf{i}_w \frac{\Delta H r_e^4}{2r^3} \quad . \quad . \quad . \quad . \quad . \quad (18)$$

where $\mathbf{A}_2 = \mathbf{i}_w(\Delta H r_e^4/2r^3)$ will generate a field falling off as r^{-4} but independent of w. Here $\Delta H = H_A - H_c$ is the difference between the observed horizontal field at the point of observation along the equator and the dipole horizontal field at that point.

Fig. 5



Allowed and forbidden regions in the R–γ plane as applied to motion in the equatorial plane of a dipole.

Equation (4) still applies since $\mathbf{A}$ is not a function of w. Introducing the Störmer unit of length $C = \sqrt{(M/P)}$ we may write eqn. (4) in the form

$$2\gamma = R \sin\theta + \frac{1}{R}\left(1 + \frac{\Delta H R_e}{2H_c R}\right) \quad . \quad . \quad . \quad . \quad . \quad (19)$$

where

$$R = \frac{r}{c} \quad R_e = \frac{r_e}{c} \quad H_c = \frac{M}{r_e^3}.$$

A method for finding the cut-off rigidity implied by an equation similar to eqn. (19) in which R_e appears explicitly has been developed by Treiman (1953). Consider the motion of the particles in a plane which has rectangular coordinates R and γ. The particle orbits are represented by straight lines of constant γ in this plane. The allowed regions are determined by the requirement that $|\sin\theta| \leqslant 1$. As in ordinary Störmer theory, the boundary line between the allowed and forbidden regions, obtained

by putting $\sin\theta = 1$ in eqn. (19), splits the $R-\gamma$ plane up into two types of allowed regions (fig. 5). Below a certain value of γ, particles can reach the earth from infinity, e.g. orbit (*b*). Above this value of γ they cannot penetrate near the earth because the orbits are interrupted by a forbidden region lying between the two types of allowed regions, e.g. orbit (*a*).

The critical value of γ is given by finding the minimum in the boundary line

$$2\gamma = R + \frac{1}{R} + \frac{\Delta H R_e}{2H_c R^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Taking $\Delta H R_e/2H_c$ to be small and differentiating and solving to first order only we find that this minimum is given by

$$R = 1 + \frac{\Delta H R_e}{2H_c}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

$$2\gamma = 2 + \frac{\Delta H R_e}{2H_c}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21a)$$

(Note that in ordinary Störmer theory the minimum is given by $\gamma = 1$ while here the corresponding value of γ depends on the particle rigidity since R_e appears in eqn. (21).) The value of R_e corresponding to this value of γ is found by substituting eqn. (21*a*) in eqn. (19) with $R = R_e$ and solving for R_e. Since we are only interested in vertical cut-off rigidity, $\sin\theta$ is put equal to zero. Then

$$R_e = \frac{1}{2}\left(1 + \frac{3}{8}\frac{\Delta H}{H_c}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

is obtained, neglecting all but first order terms. Remembering that $r_e = CR_e$, the vertical cut-off rigidity is found to be

$$P_M = \frac{M}{4r_e^2}\left(1 + \frac{3}{4}\frac{\Delta H}{H_c}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

Similar expressions are obtained when the above calculation is carried out for non-dipole fields falling off as r^{-5}, r^{-6}, r^{-7} and r^{-8} with no w dependence, the only difference being that the numerical factor $\frac{3}{4}$ is replaced by 0·58 for the r^{-5} field, by 0·47 for r^{-6} field, by 0·39 for r^{-7} field and by 0·33 for the r^{-8} field. If, as before, we weight these terms according to the average magnitude over the earth of the field contribution due to the respective harmonics, we obtain an effective value of 0·60. Thus (23) becomes

$$P_M = \frac{M}{4r_e^2}\left[1 + 0{\cdot}6\frac{\Delta H}{H_c}\right] \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

which is simply the Störmer expression with a new effective magnetic moment

$$M' = M\left(1 + 0{\cdot}6\frac{\Delta H}{H_c}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

In a region $\pm 20°$ on either side of the equator the reasoning leading to eqn. (24) remains justifiable since the particle trajectories are still nearly

at right angles to the magnetic field near the earth and motion perpendicular to the equatorial plane is at a minimum. Some additional latitude correction should be present, however ($\cos\lambda \neq 1$ in eqn. (19)). Since the last stages of the approach of the particle follow quite closely the radius through the point of arrival, a satisfactory approximation to the latitude part of the correction is obtained by fitting a dipole field to the real field out along the radius through the point of arrival. This fitting is carried out by matching $\int H\,dr$ for the two fields. Consequently we have

$$\bar{\lambda} = \tan^{-1}\left(\frac{\bar{V}}{2\bar{H}}\right). \qquad \ldots\ldots\ldots \quad (26)$$

where $\bar{V}$ and $\bar{H}$ are related to the surface field by equations (12 *a*) and (12 *b*). It will be seen that the relationship between $\bar{\lambda}$ and the surface field turns out to be the same here as at high latitudes although the underlying reasoning is rather different.

The expression for the cut-off rigidity in the equatorial region between $\pm 20°$ therefore becomes

$$P_M = \frac{M}{4r_e^2}\left[1 + 0{\cdot}6\left(\frac{H_A - H_c}{H_c}\right)\right]\cos^4\bar{\lambda}. \qquad \ldots \quad (27)$$

§ 5. Evaluation of the Cut-off Rigidities

We have deduced above two expressions for the vertical cut-off rigidity for particles in the earth's field. For latitudes greater than 30° we have

$$P_M = \frac{M}{4r_e^2}\cos^4\bar{\lambda} \qquad \ldots\ldots\ldots \quad (28)$$

and for latitudes less than 20°

$$P_M = \frac{M}{4r_e^2}\left[1 + 0{\cdot}6\left(\frac{H_A - H_c}{H_c}\right)\right]\cos^4\bar{\lambda} \qquad \ldots\ldots \quad (29)$$

where

$$\bar{\lambda} = \frac{\tan\lambda_c}{1 + \dfrac{H_A}{H_c}} + \frac{\tan\lambda_A}{1 + \dfrac{H_c}{H_A}} \qquad \ldots\ldots\ldots \quad (30)$$

λ_c is the geomagnetic latitude† and λ_A is the dip latitude ($2\tan\lambda_A = \tan\delta_A$ where δ_A is the dip angle).

---

† The system of geomagnetic coordinates used above are those derived from the 1945 magnetic survey (Jory 1956). The formulae relating geomagnetic latitude and longitude to geographic coordinates are :

$$\sin\lambda = \cos\theta_0\cos\theta\cos(w - w_0) + \sin\theta_0\sin\theta$$

$$\cos\Lambda = \frac{\sin\theta_0\cos\theta\cos(w - w_0) - \cos\theta_0\sin\theta}{\cos\lambda}$$

where λ = geomagnetic latitude, θ = geographic latitude, Λ = geomagnetic longitude, w = geographic longitude and $\theta_0 = 78{\cdot}6°$; $w_0 = 290{\cdot}0°$; corresponding to the position of the 1945 geomagnetic north pole.

It will be noted that eqn. (28) has no correction factor simply proportional to the local H as in eqn. (29). This is a direct consequence of two assumptions: (*a*) the distortion of the equatorial jaws by the non-dipole parts of the field is unimportant at high latitudes, and (*b*) the particles approach their high latitude arrival point with their main component of motion along the line of force through this point.

In the region between 20° and 40° it is not possible in any simple way to deduce an expression for the cut-off rigidity. For this region we simply use values given by the arithmetic mean of the above two expressions.

The values of the surface field parameters H_A and δ_A are taken from Admiralty charts which are constructed using the first six terms in the spherical harmonic expansion of the earth's field.

Cut-off rigidities calculated using expressions (28) and (29) are given in table 2. In addition, the differences between these modified cut-offs and the usual centred dipole cut-offs are shown in fig. 6 as a contour map.

The values of the cut-off rigidities deduced above are subject to some error because of the approximate nature of the calculations. It is not possible to give an accurate estimate of the uncertainties involved, but we can obtain a rough idea of their magnitude by considering what we believe to be the three major sources of error. These are as follows:

(*a*) In calculating the cut-off rigidities we have used average values of the non-dipole terms rather than the actual values at each point. There are two sources of error which arise as a result of this procedure:

(i) For any particular point the individual non-dipole terms may have any value between zero and the maximum possible. This means that the values of the weighting factors, 0·5 and 0·6, are not strictly constant but vary from place to place. It can be shown that the maximum variation amounts to $\pm 0{\cdot}1$.

(ii) At any point on the earth's surface where the field is identical in intensity and direction with that due to the centred dipole alone the expressions deduced above for the cut-off rigidity reduce to those for the centred dipole. In general, however, the cut-off rigidity at such a point is not that of the dipole alone because, although the sum of the non-dipole terms is zero, it does not follow that each individual term is also zero. Consequently, although the field at the earth's surface is identical to that of a dipole, it will not be so at points vertically above.

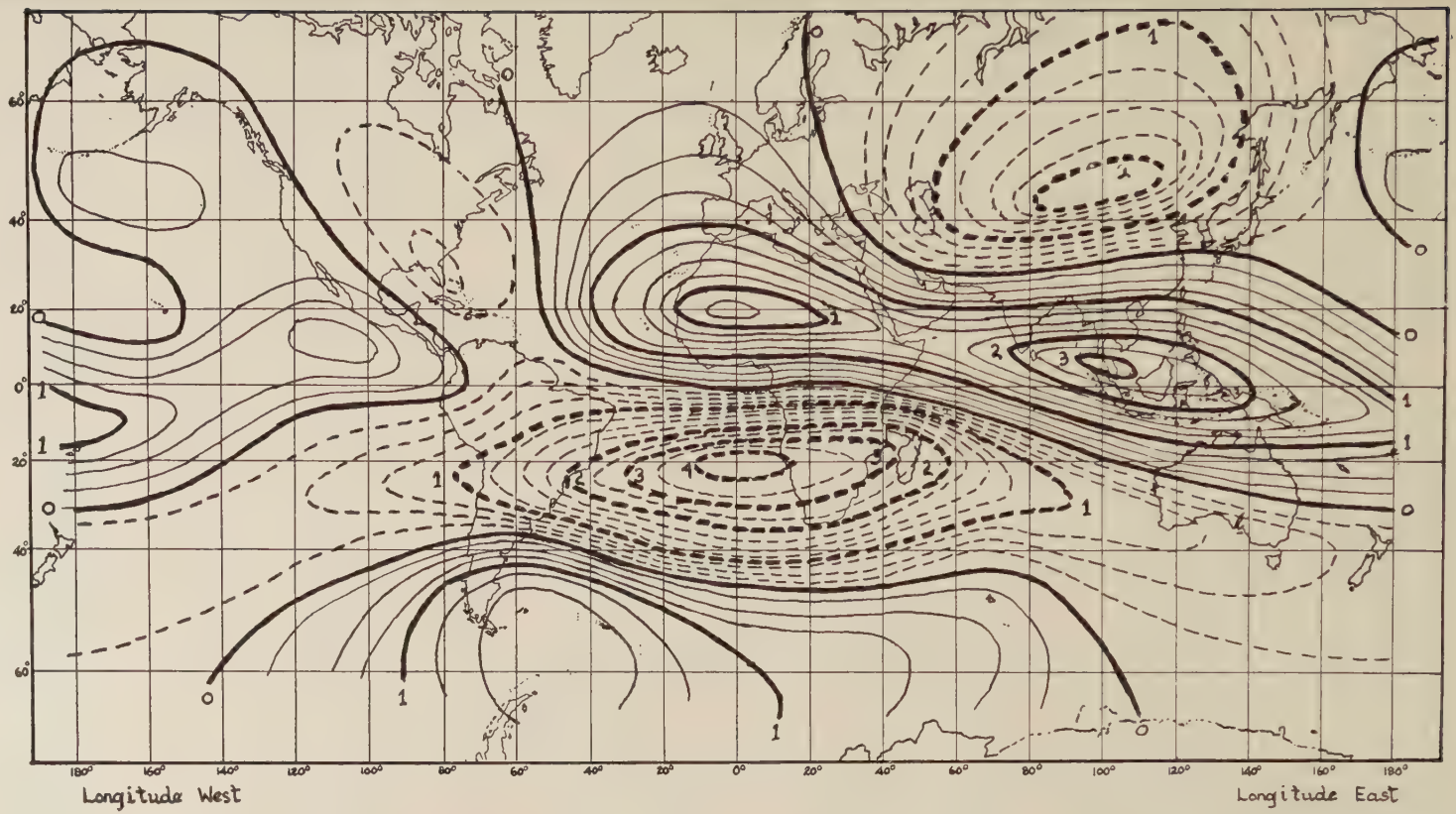
In the equatorial region between latitudes $\pm 20°$ the error due to both these causes has been determined to be $\sim 0{\cdot}4$ Bev/c. At high latitudes the error is given by $0{\cdot}2\Delta P$(Bev/c) where ΔP is the correction to the dipole cut-off rigidity plotted in fig. 6; or by $4\sin^2\lambda$ (%) whichever is the greater. The first value is that arising from the uncertainty in the weighting factor whilst the second is the result of *a*(ii).

In the intermediate region between 20° and 40° there are two further sources of error which we now discuss under (*b*) and (*c*).

(*b*) In table 2 the cut-off rigidities at geomagnetic latitude 30° are the arithmetic means of the values given by expressions (28) and (29). In

Table 2. Modified Cut-off Rigidities in the Earth's Field in Bev/c

λ \ Λ	Dipole	0	15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	240	255	270	285	300	315	330	345
+70°	0·21	0·16	0·22	0·30	0·35	0·36	0·35	0·34	0·29	0·24	0·17	0·12	0·07	0·05	0·05	0·08	0·12	0·16	0·16	0·16	0·13	0·10	0·09	0·09	0·12
+60°	0·93	0·75	0·93	1·10	1·22	1·29	1·26	1·15	1·0	0·85	0·68	0·45	0·31	0·22	0·34	0·55	0·84	1·0	1·05	1·03	0·88	0·75	0·67	0·59	0·62
+50°	2·52	2·2	2·5	2·9	3·2	3·3	3·3	3·0	2·6	2·3	1·9	1·54	1·16	1·10	1·36	2·0	2·56	2·74	2·76	2·8	2·68	2·54	2·40	2·16	2·09
+40°	5·20	4·8	5·0	5·7	6·3	6·4	6·5	5·9	5·4	4·7	4·1	3·7	3·2	3·15	3·6	4·6	5·0	5·5	5·5	5·5	5·4	5·4	5·3	5·1	4·7
+30°	8·5	8·3	8·5	9·2	10·0	10·6	10·6	10·0	9·2	8·4	7·5	6·9	6·5	6·7	7·4	8·0	8·4	8·5	8·5	8·5	8·6	8·9	9·0	8·8	8·3
+20°	11·	11·6	11·5	12·2	12·9	13·8	14·2	14·1	13·6	12·6	11·8	11·6	11·7	11·8	11·8	11·8	11·7	11·5	11·6	11·6	12·0	12·4	12·7	12·6	12·1
+10°	14·2	13·7	13·5	13·7	14·2	14·6	15·1	15·7	16·0	15·9	15·8	15·9	16·1	15·7	15·2	14·9	14·5	14·2	14·0	14·3	14·6	14·7	14·7	14·7	14·3
0°	15·0	14·2	13·8	13·6	13·3	13·1	13·4	14·2	15·3	16·2	17·0	17·6	18·0	17·8	17·1	16·5	16·2	16·0	15·8	15·7	15·6	15·4	15·0	14·9	14·6
−10°	14·2	13·1	12·5	11·9	11·0	10·2	10·1	10·6	12·1	13·6	15·0	16·0	16·6	16·9	16·8	16·3	15·9	15·7	15·4	15·0	14·7	14·1	13·8	13·5	13·3
−20°	11·7	10·9	10·6	9·6	8·6	8·1	8·0	8·5	9·5	9·5	10·9	11·6	12·6	13·2	13·7	13·8	13·2	13·0	12·8	12·3	11·9	11·5	11·3	11·2	11·1
−30°	8·5	9·1	9·1	8·2	7·3	6·4	5·6	5·5	5·8	6·6	7·2	7·4	7·8	8·5	8·8	8·6	8·8	8·6	8·5	8·5	8·4	8·3	8·2	8·4	8·9
−40°	5·20	6·8	7·0	6·4	5·7	4·9	4·2	4·2	4·5	4·6	4·5	4·2	4·1	4·4	4·7	4·7	4·7	4·7	4·8	4·7	4·8	5·0	5·1	5·6	6·2
−50°	2·52	4·2	4·3	4·1	3·7	3·3	3·0	3·0	2·9	2·8	2·50	2·21	1·92	1·78	1·84	1·96	2·15	2·20	2·20	2·21	2·24	2·31	2·56	3·0	3·6
−60°	0·93	2·08	2·25	2·27	2·14	2·00	1·90	1·82	1·70	1·58	1·46	1·05	0·78	0·62	0·49	0·45	0·48	0·51	0·62	0·70	0·73	0·84	1·02	1·33	1·70
−70°	0·21	0·74	0·86	0·93	0·95	0·93	0·93	0·91	0·87	0·74	0·60	0·44	0·28	0·17	0·09	0·06	0·05	0·06	0·08	0·08	0·13	0·18	0·27	0·42	0·58



Contour map showing differences between centred dipole cut-off rigidities and the modified cut-off rigidities derived in this paper. Positive differences are shown as solid lines negative differences as dashed lines. Contour intervals are 0·25 BEV/c from 0 to 2 BEV/c and 0·5 BEV/c above 2 BEV/c. Heavy lines occur at 1 BEV/c intervals.

general these values do not differ by more than 0·5 BeV/c and consequently the error due to this procedure is not greater than 0·25 BeV/c. The only exceptions are the values at longitudes 45, 60, 75 and 90, corresponding to points lying in the South African anomaly, where the error may be as much as 1 BeV/c.

(*c*) At latitudes between 15° and 30°, because of the penumbra, the Störmer expression (7) does not give the correct cut-off rigidities even for a simple dipole field. Indeed at these latitudes there is no unique cut-off rigidity and it is necessary to compute individual trajectories in order to know whether they are allowed or forbidden. (Strictly this is true at all latitudes but outside the limits given above the penumbra may be neglected without introducing appreciable error.) Approximate corrections to the Störmer cut-offs due to the penumbra have been given by a number of authors (Lemaitre and Vallarta 1936, Schwartz 1956) on the assumption of a purely dipole field, but if we take into account the non-dipole terms these corrections will, of course, be changed. The best approach to the problem would seem to be an experimental one such as was adopted, for example, by Neher (1952) who compared intensities at different directions in the E–W plane at the equator with those at an equivalent depth at different latitudes. When re-interpreted in terms of the cut-offs derived in this paper, these data show that a correction should still be applied for the penumbra although it is somewhat smaller than that deduced theoretically for the simple dipole. By detailed comparison of the latitude and longitude variations of the nucleonic component at aeroplane altitudes in the equatorial region it should be possible to deduce an adequate correction for the penumbra from experimental data already in existence. This correction may amount to as much as 1 BeV/c in the region 20° to 30° geomagnetic latitude and consequently the values of the cut-off rigidity in this region should be treated with reserve.

§ 6. Comparison of Cut-off Rigidities with Experimental Results

The experimental data which are of sufficient accuracy to indicate departures from the usual dipole theories include (1) the distribution of the additional isotropic radiation over the earth's surface during the flare of February 23, 1956, (2) measured cut-off rigidities for alpha particles in North America and Europe and (3) the position of the cosmic ray equator as determined by the aeroplane survey of Simpson.

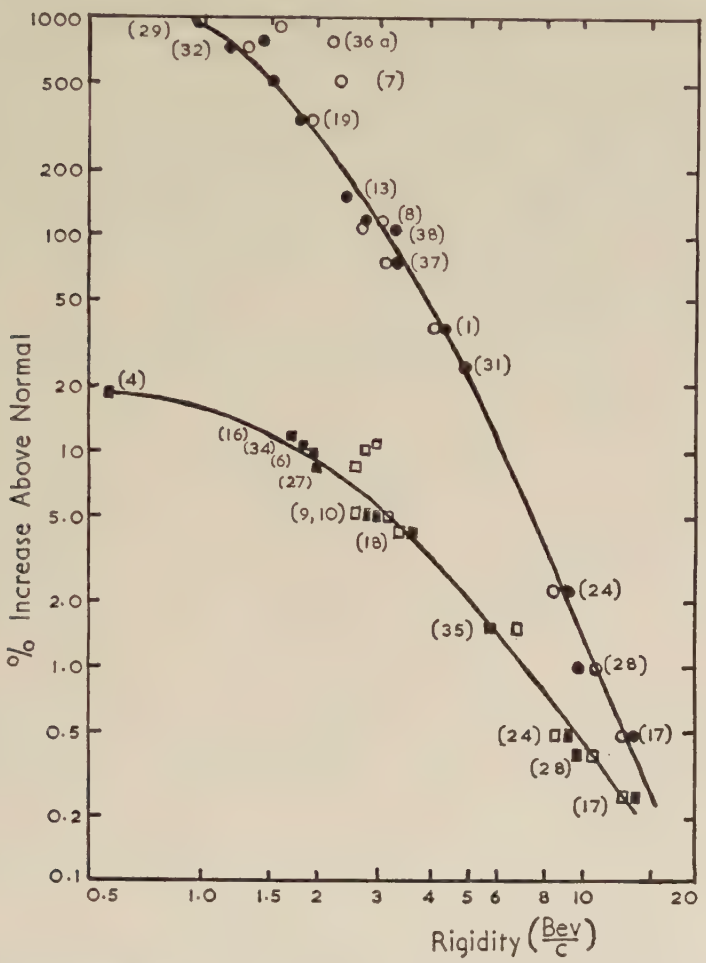
Each of these groups of data will now be discussed separately.

6.1. *Solar Flare of February* 23, 1956

Subject to the assumption that the flare radiation arriving after approximately 5.00 G.M.T. was isotropic this event provides perhaps the most sensitive measure available of the relative cut-off rigidities for the stations at which the event was observed. This is so because of the extremely

steep rigidity spectrum of the flare particles which produced a latitude effect of an order of magnitude greater than normal.

Fig. 7



Percentage increase above normal at 6·00 G.M.T. for the flare of February 23, 1956 for neutron monitors and ion chambers with ~10 cm Pb, plotted as a function of cut-off rigidity. Numbers refer to stations listed in the Appendix.

Neutrons	●	Modified rigidities.
	○	Eccentric dipole rigidities.
Ion chambers	■	Modified rigidities.
	□	Eccentric dipole rigidities.

In fig. 7 are shown the percentage increases above normal at 6.00 G.M.T. for stations using (A) neutron monitors and (B) ion chambers with 10 cm Pb (Gold and Elliot 1956). In the case of certain low latitude stations the relative increases are obtained for 5.00 and 5.30 G.M.T. and normalized

to mid-latitude increases at 6.00 G.M.T. in order to improve the accuracy of the data. All data have been reduced to a constant atmospheric depth (1030 g/cm^2). In order to do this we have assumed that the rigidity spectrum of the flare particles is so steep that the main contribution to the increases recorded in the lower atmosphere is due to particles just above cut-off rigidity. The details of the method used will be discussed in a later paper.

In the figure the solid symbols refer to the modified cut-off rigidities derived in this paper and the open symbols to those derived from the latest eccentric dipole model. It is clear that the modified cut-off rigidities provide a more consistant picture of the flare increase than the eccentric dipole cut-offs, e.g. increases at North American and European stations. In fact the increases at all stations now lie in proper order when plotted against the modified cut-off rigidities.

The root mean squared deviations in per cent of the expected cut-off rigidity as predicted from the respective best fit flare response curves are shown in table 3.

Table 3

	Eccentric dipole	Modified dipole
Neutron stations (14)	12·6%	7·5%
Ion chamber stations (12)	14·3%	5·3%

6.2. *Measured Cut-off Rigidities*

In table 4 are compared cut-off rigidities for alpha particles measured at various points in North America and Europe and the values calculated according to our method. The corresponding values for the centred and eccentric dipole model are also shown.

Table 4. Comparison of Measured and Predicted Cut-off Rigidities

Place		Cut-off Rigidity (BEV/c) P_c	P_E	P_M	P_{meas}
Minneapolis	(1)	1·55	1·74	1·10	$<1·0$
$\lambda=54·3, \Lambda=335$	(2)	1·70	1·91	1·23	$1·15\pm0·1$
Waukon	(3)	1·78	1·97	1·33	$1·2\ \pm0·1$
$\lambda=53·0, \Lambda=334$	(2)	1·93	2·15	1·45	$1·4\ \pm0·1$
N. Missouri	(3)	2·35	2·61	1·89	$1·8\ \pm0·1$
Bristol	(4)	1·65	2·05	2·10	$2·4\ \pm0·1$
N. Italy	(5)	3·50	3·85	4·0	$4·6\ \pm0·2$
Sardinia	(6)	5·05	5·43	5·45	$5·6\ \pm0·4$

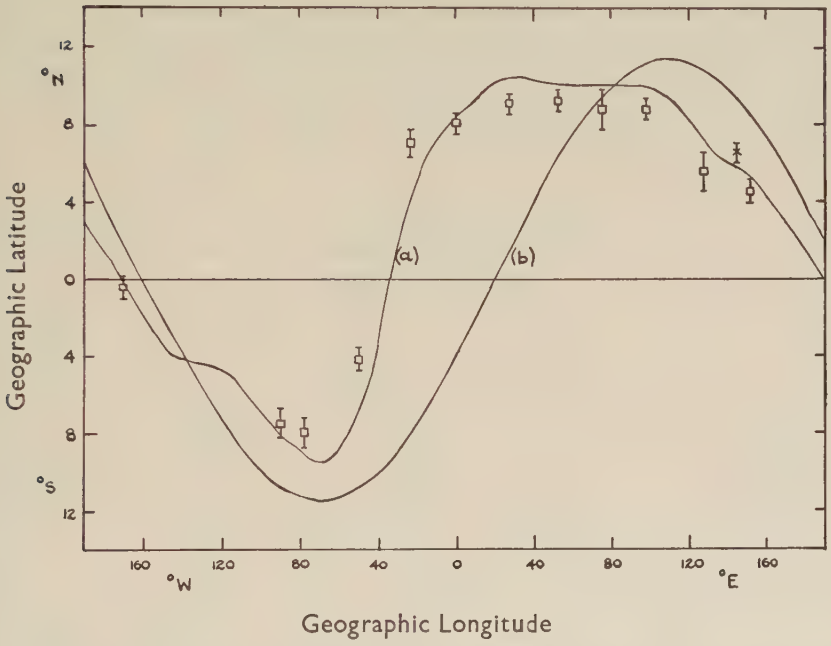
(1) Fowler *et al.* (1957), (2) McDonald (1957), (3) Freier *et al.* (1958), (4) Waddington (1956), (5) Fowler and Waddington (1956), (6) Aly and Waddington (1957).

It is clear from the table that the modified cut-off rigidities agree much more closely with the measured values than do the values calculated using the eccentric or centred dipole. Significant discrepancies still remain, however, between the measured cut-offs at Bristol and N. Italy and those predicted using the modified dipole.

6.3. *The Cosmic Ray Equator*

Simpson (1957) has recently determined the position of minimum cosmic ray intensity by a series of aeroplane flights at different longitudes in the equatorial region. In fig. 8 we show the line of maximum cut-off

Fig. 8



Comparison of the cosmic ray equator measured at aeroplane altitudes. (Simpson (1957) □, Storey *et al.* (1958) ×) with the line of minimum intensity predicted using the modified cut-off rigidities (*a*), and the geomagnetic equator (*b*).

rigidity as a function of geographic longitude, calculated using the modified cut-off equations developed above. The positions of the cosmic ray intensity minima found by Simpson are plotted on the figure and the geomagnetic equator is also drawn in. It is seen that the experimental points all lie within about 2° of the positions predicted from the calculated cut-offs.

6.4. *The Aurora*

In addition to the experimental cosmic ray evidence, auroral observations at high latitudes may be used to check the theory developed above. It is generally thought that the aurora is caused by low energy charged particles entering the earth's atmosphere at high latitudes along magnetic lines of force. Because of this one would expect the auroral zone to be roughly concentric about the geomagnetic pole and exactly so if we are dealing with a purely dipole field. We have shown, however, that for cosmic rays in the earth's field $\bar{\lambda}$ is a more satisfactory parameter than the geomagnetic latitude. As a consequence we might reasonably expect the auroral zone to lie along lines of constant $\bar{\lambda}$ which are not necessarily concentric about any of the geomagnetic poles.

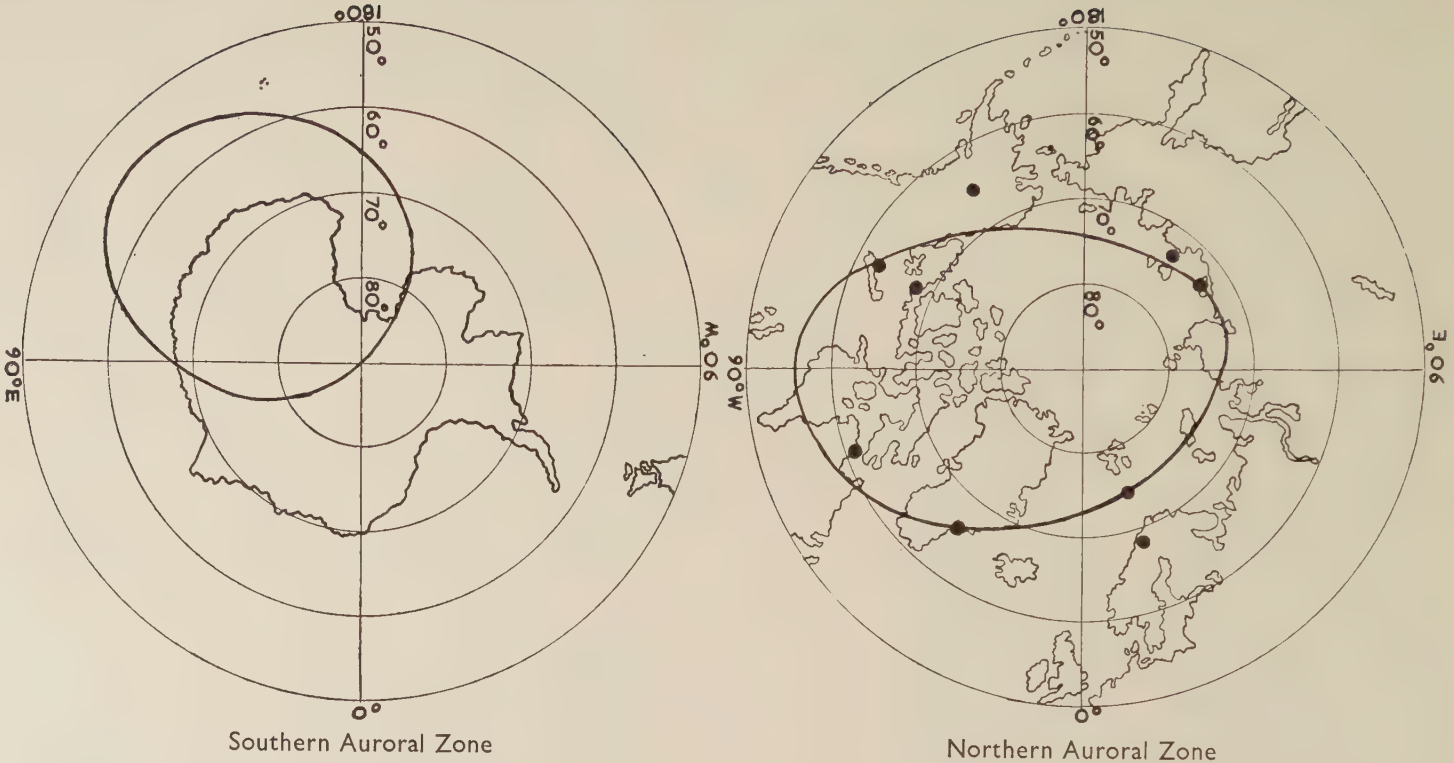
Table 5 shows nine observing stations in the northern hemisphere operating during the sunspot minimum period 1932 33. These stations, selected from Vestine's (1944) list, recorded an aurora on all nights when observations were possible and are classified by Vestine as 100% stations. In the table are also listed the corresponding geomagnetic latitudes, dip latitudes and values of $\bar{\lambda}$. Mean values of these parameters have been formed and also the sum of the squares of their residuals.

Table 5. Effective Latitudes, Geomagnetic Latitudes and Dip Latitudes of Auroral Stations

Place	λ (geomagnetic	λ (effective)	λ_A (dip)
Bear Island	71·1	70·6	70·2
Ssagastyr	62·2	69·8	77·6
Abisko	66·0	65·6	65·2
Coppermine	73·7	76·9	79·9
Angmagsalik	74·2	70·9	67·2
Fort Rae	69·0	72·2	75·1
Cape Hope's Advance	72·6	73·4	74·1
Great Liakhovsky Island	62·8	69·0	75·2
College—Fairbanks	64·7	64·8	64·9
	mean = 68·5	mean = 70·4	mean = 72·2
	$\sum(\Delta\lambda)^2 = 176{\cdot}3$	$\sum(\Delta\bar{\lambda})^2 = 111{\cdot}6$	$\sum(\Delta\lambda_A)^2 = 240{\cdot}8$

It will be seen that the stations are better represented by a constant $\bar{\lambda}$ than by either a constant geomagnetic latitude or dip latitude. The mean value of $\bar{\lambda}$ is 70·4°. In fig. 9 the line $\bar{\lambda} = 70{\cdot}4°$ has been plotted together with the stations from table 5. In fig. 9 we have also plotted the line $\bar{\lambda} = 70{\cdot}4°$ on a map of the southern polar region corresponding to where we would expect the southern auroral zone to lie at sunspot minimum.

Fig. 9



Predicted northern and southern zones of maximum auroral frequency. Northern auroral stations with 100% frequency are shown as solid points.

Observational data in the Antarctic have, however, been too sparse to permit any very reliable estimate of the position of the southern auroral zone.

In addition to the experimental evidence discussed above, there are other data which are difficult to explain in detail using the usual dipole cut-off rigidities, particularly the latitude survey data, obtained by various authors, for the neutron and ionizing components at aeroplane altitudes and at sea-level. These data will be considered in a future paper dealing with the differential rigidity response curves of the various secondary components.

Acknowledgments

We wish to acknowledge the generous collaboration of Dr. H. Elliot during the course of the work reported here. We also thank Mr. R. J. Hynds who integrated the particle trajectory shown in fig. 2.

Both of the authors are indebted to the Department of Scientific and Industrial Research for financial support.

APPENDIX

Calculated Cut-off Rigidities for Selected Cosmic Ray Stations

	Station	P_E	P_M		Station	P_E	P_M
1	Albuquerque	4·1	4·3	21	Macquarie Is.	0·54	0·50
2	Alma Ata	7·8	5·8	22	Makerere	14·3	14·9
3	Berkeley	3·9	4·4	23	Mawson	0·14	0·75
4	Cape Schmidt	0·46	0·56	24	Mexico City	8·5	9·1
5	Chacaltaya	13·0	13·6	25	Mina Aquilar	12·1	13·1
6	Cheltenham	2·84	1·96	26	Minneapolis	1·74	1·10
7	Chicago	2·22	1·54	27	Moscow	2·52	1·97
8	Climax	3·0	2·71	28	Mt. Norikura	10·8	9·6
9	Christchurch	2·50	2·79	29	Ottawa	1·60	0·96
10	Freiburg	3·0	3·0	30	Pic-du-Midi	3·9	4·2
11	Fort Churchill	0·29	0·11	31	Sacremento Pk.	4·8	4·7
12	Godhavn	0·013	0·04	32	Stockholm	1·33	1·19
13	Gottingen	2·42	2·42	33	Sulpher Mtn.	1·12	0·98
14	Hawaii	12·4	11·3	34	Sverdlovsk	3·0	1·86
15	Hermanus	7·1	4·8	35	Tiblisi	6·6	5·7
16	Hobart	1·71	1·70	36	Mt. Washington	1·85	1·03
17	Huancayo	13·2	14·3	36a	Durham	2·15	1·42
18	Innsbruck	3·4	3·5	37	Weissenau	3·2	3·3
19	Leeds	1·84	1·77	38	Wellington	2·70	3·3
20	London	2·15	2·25	39	Mt. Wrangell	0·48	0·63
				40	Yakutsk	2·05	1·25

References

Aly, P., and Waddington, C. J., 1957, *Nuovo Cim.*, Ser. X, **5**, 1679.

Fermi, E., 1949, *Nuclear Physics* (Chicago : University Press), p. 224.

Finch, H. P., and Leaton, B. R., 1957, *M.N.R.A.S. Geophysical Suppl.*, **1**, 314.

Fowler, P. H., and Waddington, C. J., 1956, *Phil. Mag.*, Ser. VIII, **1**, 637.

Fowler, P. H., Waddington, C. J., Freier, P. S., Naugle, J., and Ney, E. P., 1957, *Phil. Mag.*, Ser. VIII, **2**, 157.

Freir, P. S., Ney, E. P., Fowler, P. H., 1958, *Nature, Lond.*, **181**, 1319 ; and *Proc. Varenna Conference* (to be published in *Nuovo Cimento*).

Gold, T., and Elliot, H., 1956, data collected in "The Solar Cosmic Ray Outburst 1956 February 23".

Johnson, T. H., 1938, *Rev. mod. Phys.*, **10**, 193.

Jory, F. S., 1956, *Phys. Rev.*, **102**, 1167.

Kodama, M., and Miyazaki, Y., 1957, *Rep. of Ionosphere Res. in Japan*, **11**, 99.

Lemaitre, G., and Vallarta, M. S., 1936, *Phys. Rev.*, **50**, 493.

McDonald, F. B., 1957, *Phys. Rev.*, **107**, 1386.

Neher, H. V., 1952, *Prog. in C.R. Physics*, Vol. I (Amsterdam : North Holland Publ. Co.), p. 243.

Pfotzer, G., 1957, *Proc. Varenna Conference* (to be published in *Nuovo Cimento*).

Rose, D. C., Fenton, K. B., Katzman, J., and Simpson, J. A., 1956, *Canad. J. Phys.*, **34**, 968.

Rothwell, P., and Quenby, J. J., 1957, *Proc. Varenna Conference* (to be published in *Nuovo Cimento*).

Rothwell, P., 1958, *Phil. Mag.* (to be published).

Schmidt, A., 1934, *Beit. Angew. Geophys.*, **41**, 346.

Schwartz, M., 1956, *Bull. Amer. phys. Soc.*, Ser. II, **1**, 319.

Simpson, J. A., Fenton, K. B., Katzman, J., and Rose, D. C., 1956, *Phys. Rev.*, **102**, 1648.

Simpson, J. A., 1957, *Proc. Varenna Conference* (to be published in *Nuovo Cimento*).

Skorka, S., 1958, *Z. Phys.*, **151**, 630.

Storey, J. R., Fenton, A. G., and McCracken, K. G., 1958, *Nature, Lond.*, **181**, 1153.

Störmer, C., 1955, *The Polar Aurora* (Oxford : University Press).

Treiman, S. B., 1953, *Phys. Rev.*, **89**, 130.

Vestine, E. H., 1944, *Journal Ter. Mag. and Atm. Elect.*, **49**, 77.

Waddington, C. J., 1956, *Nuovo Cim.* Ser. X, **3**, 930.

The Photo Creation and Destruction of F Centres†

By R. V. Hesketh
University of Glasgow

[Received October 16, 1958]

Abstract

F centres formed within, and not at external surfaces of KCl crystals have been observed. The rate of creation by x-rays has been found to be of the form $dN/dt = A - BN$, where A and B are temperature dependent constants. The constant A has been found to be close to the rate of liberation of electrons multiplied by an energy factor $\exp(-0{\cdot}05\ \text{ev}/kT)$. It is suggested the F centres are formed from electrons trapped at incipient negative ion vacancies. Optical bleaching shows that the thermal activation process from the excited state to the conduction band is similar to that in additively coloured crystals. The quantum efficiency of bleaching indicates that less than 10^{15} per cm^3 of frozen-in vacancies take part in colour centre formation and that at 300°K the created vacancies are only stable when occupied by electrons.

§ 1. Introduction

When an alkali halide crystal is exposed to ionizing radiation, electrons and holes are released within the crystal and may eventually be trapped at ion vacancies to form colour centres. It is known that the energy required to form an F centre is not simply that required to release an electron from a negative ion, and estimates for formation by u.v., x, and γ-rays vary from this energy to two hundred times this energy (Seitz 1954, p. 63). From the rate of coloration it is known that the energy required increases as the temperature is lowered (Harten 1949). The present paper gives some quantitative results on the rate of coloration and its temperature variation.

Colour centres formed by ionizing radiation may be destroyed by the absorption of optical quanta. In all the present experiments it has been possible to produce complete optical bleaching, on many successive occasions, without any additional treatment of the crystal, such as annealing, which is often used to promote complete bleaching. The variation of the quantum efficiency of bleaching has been measured as a function of temperature and F centre concentration, and a qualitative interpretation of the results is given.

§ 2. Experimental Details

The crystals used in these experiments have been cleaved from a large single crystal of KCl, and have approximate dimensions $1 \times 1 \times 1{\cdot}5$ cm.

† Communicated by the Author.

They have been mounted in the cryostat shown in fig. 1. Within the cryostat are two radiation shields, the outer one having quartz windows to ensure that thermal radiation to the crystal is a minimum and that temperature gradients within it are avoided. In cooling the crystal from

Fig. 1

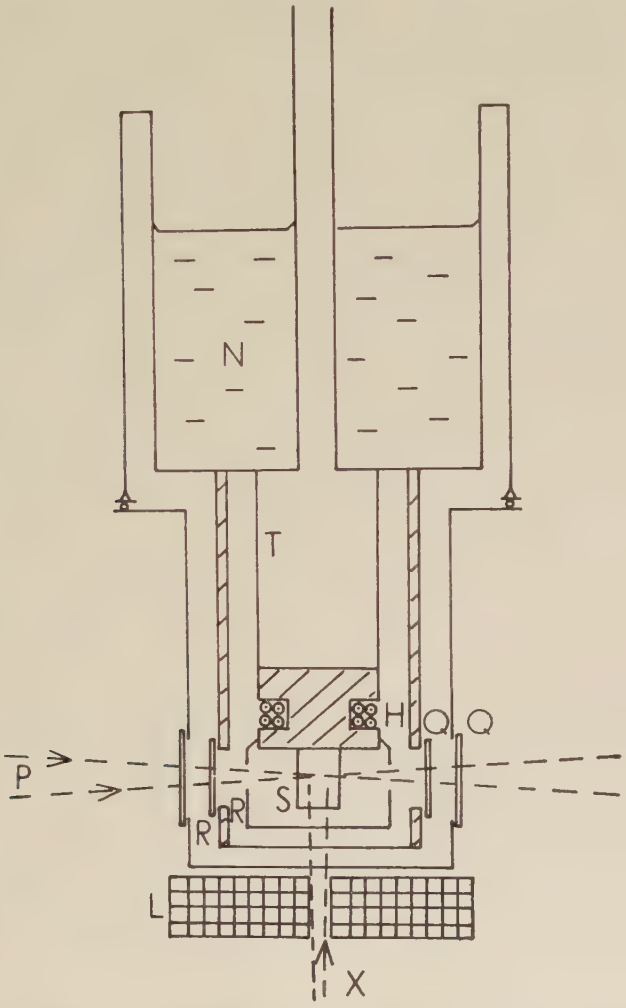


Diagram of cryostat. H, heating coil ; L, lead shield ; N, liquid nitrogen ; P, spectrophotometer beam ; Q, quartz windows ; R, radiation shields ; S, crystal ; T, stainless steel tube ; X, x-ray beam.

290°K liquid nitrogen refrigerant is placed in the outer jacket only, and the crystal cooled slowly through the greater part of the temperature range, avoiding steep temperature gradients since these may fracture the crystal (Duerig and Mador 1952). If the crystal is to be at the temperature

of the refrigerant, the tube T is then also filled. Alternatively the crystal may be maintained at a steady temperature in the range 78°–350°K by dissipating up to 15 watts in the heating coil H. The temperature is measured by a copper–constantan thermocouple cemented to the crystal close to the part optically examined. The temperature stability is adequate, being better than 1°K over several hours.

The X irradiation is from a self-rectified Machlett Aeromax 12C tube with a tungsten target. It is operated at 90 pkv and 1–30 mA. The current is kept constant by a manually controlled potentiometer. A lead screen confines the x-ray beam between two planes 3 mm apart, and the parts of the crystal beyond these planes are not irradiated. Measurements have been made at a depth of 2 mm below the face on which the x-rays are incident, and over a further depth of only 1 mm, so that the softer x-rays are filtered out and the coloration is substantially uniform. The crystal is situated 10 cm from the target of the x-ray tube and the x-ray intensity at the site of the crystal has been measured with a capacitor type dosimeter.

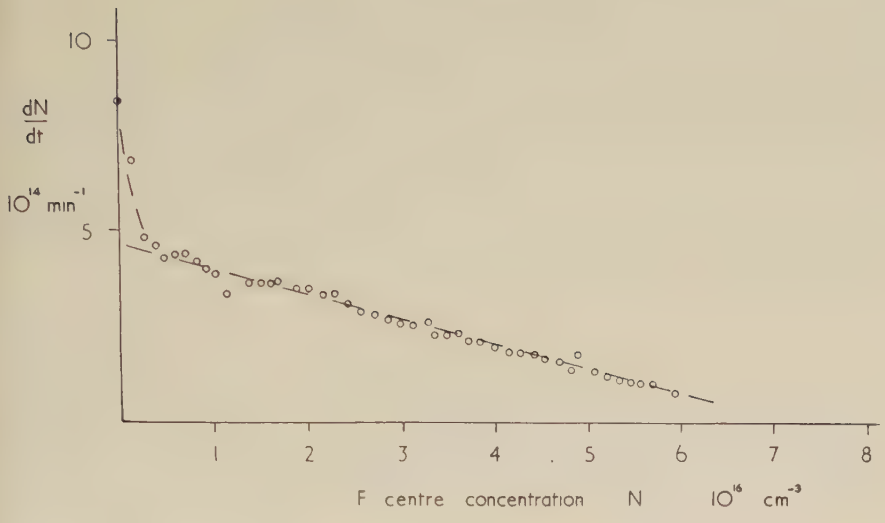
The F centre concentration has been measured by an alternating beam spectrophotometer (Hesketh 1958), the beam of which is perpendicular to the plane of the x-rays. The beam is limited and lies entirely within the crystal face on which it is incident. Thus the observations are made only on centres within the crystal and not on those at external surfaces. This is done because differences in behaviour in the two cases are known to exist (Gordon and Nowick 1956). Though Smakula's equation is not strictly applicable to the F band, it may be used to calculate the F centre concentration if the appropriate value of the oscillator strength is used. One then has $N = K \,.\, D/d$. where N is the F centre concentration, K a constant, D the optical density, $\log_{10} I_0/I_T$, and d the thickness of the coloured region.

Since 10^{13} quanta per second are used in optical bleaching and 10^8–10^9 in measurement, there is no danger of the spectrophotometer beam affecting the observations by destroying F centres. The light intensity is approximately known from the photocell and amplifier constants and indicates that the quantum efficiency of bleaching of small concentrations of F centres is close to one at room temperature. The results have been computed assuming the quantum efficiency to be precisely the oscillator strength at 303°K, a temperature a little below that at which thermal bleaching becomes significant.

§ 3. The Creation of F Centres by X-rays

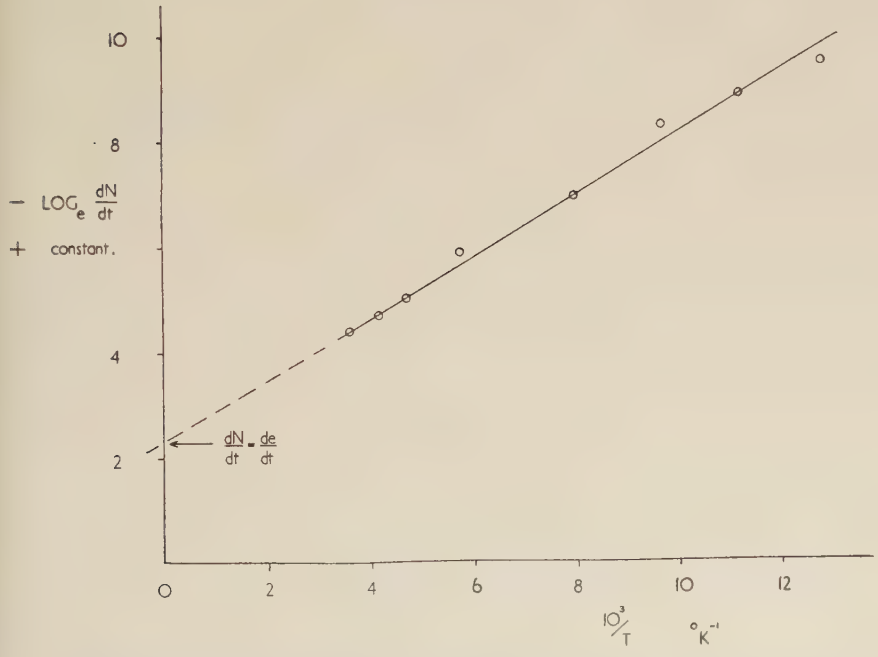
Figure 2 shows the experimental result for coloration at 285°K. An identical result has been found for different specimens cleaved from the one large crystal and for successive colorations after optical bleaching. The rate of coloration is strictly proportional to the x-ray intensity over a tube current range of 1–6 mA, and brief measurements show that the initial rate of coloration is proportional when the tube current is 30 mA. The initial, non-linear part of the curve has not been further studied, and

Fig. 2



The rate of F centre formation versus F centre concentration, at 285°K and for constant x-ray intensity.

Fig. 3



The temperature variation of the initial rate of F centre formation.

has only been apparent in colorations above 200°K. The initial rate of coloration has been measured in the range 78°–290°K, and is plotted against temperature in fig. 3. (Above 200°K the initial rate shown is that obtained by extrapolating the straight line section of fig. 2.) By extrapolation of fig. 3 a limiting value of the initial rate of coloration is found as the temperature increases indefinitely. It has been compared with the rate of liberation of electrons on the assumption that the total energy of the X-rays is used in photoionization. Agreement within a factor of 3 is found, which is good in view of the necessary approximations in the calculation. Thus from figs. 2 and 3 the following expression is derived:

$$dN/dt = A - BN. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where

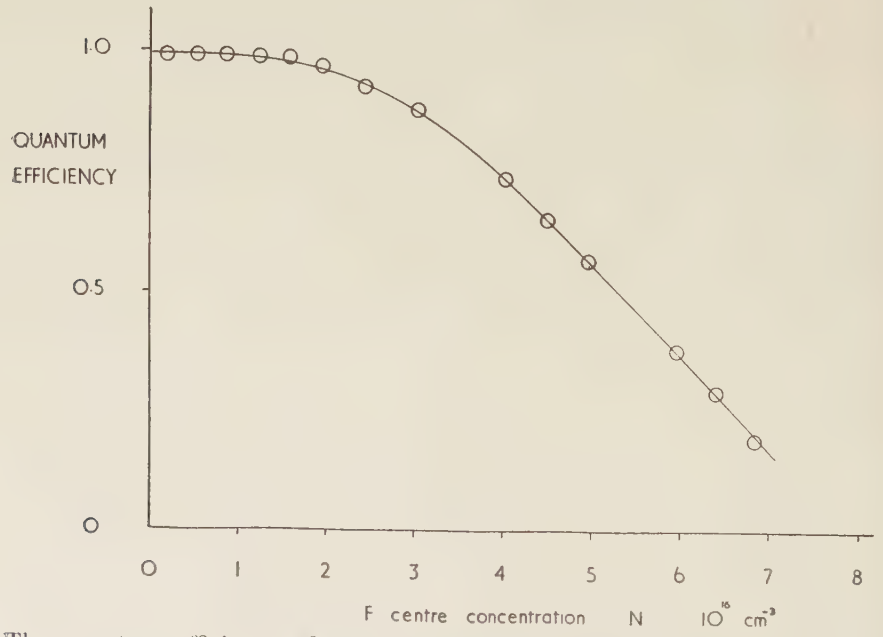
$$A = de/dt\ C \exp(-E/kT), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

de/dt is the rate of liberation of electrons, C a constant of the order of one and $E = 0{\cdot}05 \pm 0{\cdot}005$ ev. B is known to decrease as the temperature is lowered, but the temperature dependence has not yet been measured.

§ 4. The Destruction of F Centres by Optical Excitation

Over the volume of the crystal examined by the spectrophotometer beam, the X-rays create a uniform concentration of colour centres, and the rate at which these centres are optically bleached can only be simply

Fig. 4



The quantum efficiency of optical bleaching versus F centre concentration, for centres formed at 285°K.

treated while the concentration remains uniform. In the non-uniform case, after a finite amount of bleaching, the analysis becomes involved (Herman and Wallis 1955) and the experimental observations have therefore been restricted to the initial rate of optical bleaching. (The tangent at the origin has been drawn to the curve obtained by irradiation in the peak of the F band with 5×10^{14} quanta.) In each case the optical bleaching has then been completed before the next X irradiation.

The variation of $dN\ dt_{t\ 0}$ with N is shown in fig. 4. The ordinate is the quantum efficiency of bleaching in terms of the oscillator strength. It is important to notice that these observations are reproducible in successive cycles of X irradiation and bleaching. So far as this set of observations is concerned X irradiation produces no irreversible change within the crystal.

In the range where the quantum efficiency shows little variation with N (i.e. below 2×10^{16} centres cm^{-3}) its temperature variation has been measured. This is shown in fig. 5 plotted by an expression which is derived in the next section. (F′ centres formed by the optical bleaching have been destroyed by vigorous irradiation at 7500 Å before measurement of the optical density of the bleached F band.)

§ 5. Discussion

The optical bleaching results are discussed first. If light of the frequency of the F band maximum and of intensity I_0 is incident on the crystal, and and intensity I_T is transmitted, then the rate at which electrons are raised to the excited state is $f(I_0 - I_T)$, where f is the oscillator strength of the transition. The probability of thermal ejection from the excited state to the conduction band is of the form $G \exp(-E/kT)\,dt$, and the probability of return to the ground state $H\,dt$, where G and H are constants. Thus the rate at which electrons are released to the conduction band is

$$\frac{f(I_0 - I_T)}{1 + (H/G)\exp(E/kT)}.$$

From the conduction band the electrons may rejoin a positive hole or fall into a negative ion vacancy to form again an F centre. Other traps will be ignored, in which case the probability of an electron rejoining a positive hole is

$$P = \frac{\sigma_p p}{\sigma_p p + \sigma_v}(v - N)$$

where $\sigma_p \sigma_v$ are the capture cross sections and p and v local concentrations of positive holes and negative ion vacancies respectively. If the concentration of conduction electrons is small compared with N, a condition which applies during the bleaching measurements, $p \approx N$ and

$$P = \frac{\sigma_p N}{\sigma_p N + \sigma_v (v - N)}.$$

Thus the rate of bleaching (the rate of recombination of positive holes and electrons) is

$$\frac{dN}{dt} = \frac{f(I_0 - I_T)}{1 + (H/G \exp(E/kT)} \cdot \frac{\sigma_p N}{\sigma_p N + \sigma_v(v - N)}.$$

Substituting the equation $N = KD/d$, the observed rate of change of optical density is

$$\frac{dD}{dt} = \frac{fI_0(1 - 10^{-D})}{K(1 + (H/G)\exp(E/kT))(1 + (\sigma_v/\sigma_p)(v/N - 1))} \cdot \quad . \quad . \quad . \quad (3)$$

In terms of this equation the quantum efficiency of bleaching is

$$f/(1 + (H/G)\exp(E/kT))(1 + (\sigma_v/\sigma_p)(v/N - 1)).$$

At room temperature the temperature dependent term may be placed equal to one, so that fig. 4 represents the variation with N of the factor

$$f/(1 + (\sigma_v/\sigma_p)(v/N - 1)).$$

If F centres were formed simply at vacancies frozen-in during the growth of the crystal, it might be expected that saturation would occur when all the vacancies were filled, in which case the variation of quantum efficiency would be as

$$f/(1 + (\sigma_v/\sigma_p)(D_{\text{sat}}/D - 1)).$$

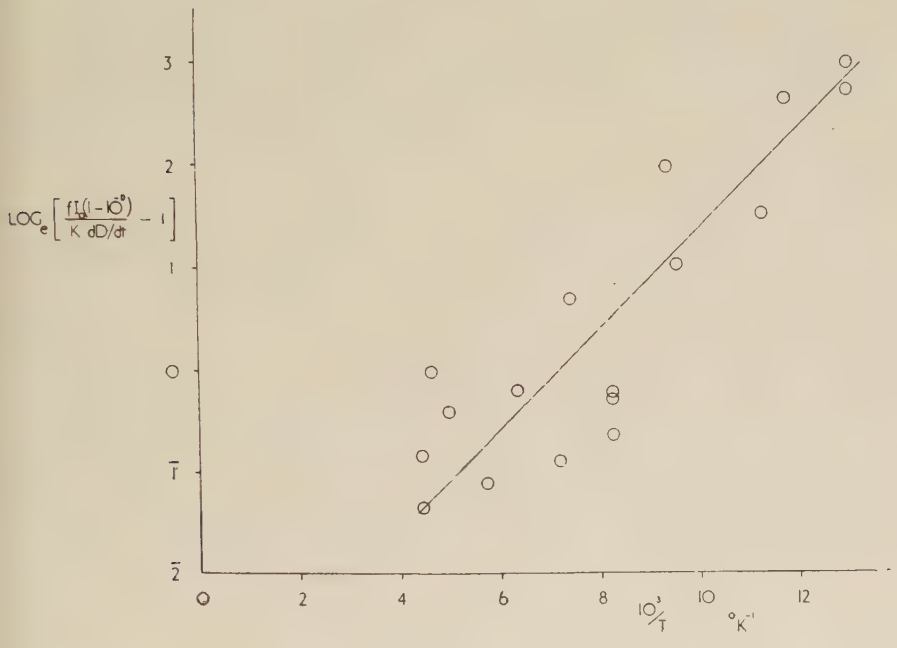
Whatever the value of σ_v/σ_p, the quantum efficiency would always be zero at $D = 0$, and always f at $D = D_{\text{sat}}$. Its form between these points would be governed by the value of σ_v/σ_p but the slope would never be negative as in fig. 4. Thus the experimental result is incompatible with the formation of F centres simply at frozen in vacancies.

There are two possibilities to explain the high quantum efficiency at low concentrations, either σ_v/σ_p is negligible, or v is equal to N; though at saturation it is necessary that σ_v/σ_p is appreciable and that v exceeds N. Positive holes and negative ion vacancies are of identical charge and similar size, so that if each remains unassociated with charged defects σ_v/σ_p should be of the order of one. If σ_v/σ_p is negligible, the negative ion vacancies must be trapped at more highly charged traps than the positive holes, and if the positive holes are trapped as V centres it is difficult to see how this could be. It might occur if the V centres are not the models proposed by Seitz, and if there is a high degree of association between ion vacancies, though this association is doubted by Theimer (1958).

On the other hand, v may be equal to N if the negative ion vacancy disappears after the release of an electron from an F centre. The disappearance of the vacancy as a single unit could be caused by the clustering of vacancies, but in view of the evidence of volume expansion of crystals during X irradiation this hypothesis is discarded in favour of the destruction of vacancies at sinks which are also the sources for creation during X irradiation. Evidence is provided by the results shown in fig. 2. Four results have

been obtained for the first irradiation of crystals, two at room temperature and two at liquid nitrogen temperature, and these are not displaced from the plot for second and latter irradiations. Thus it appears that the mechanism of vacancy production is the same in both initial and subsequent X irradiations. Since the results of fig. 4 are also reproducible, it has been concluded that the process of vacancy creation is reversible. This could be directly established if experiments of the type of Sakaguchi and Suita (quoted by Seitz 1954, p. 74) were made during the destruction of vacancies by optical bleaching. The experiments of these authors have shown that it is not necessary to suppose that a large part of the initial colouring takes place at vacancies already in the crystal, and justify the neglect of such vacancies in placing $v = N$ at low concentrations.

Fig. 5



The temperature variation of the quantum efficiency of bleaching of F centres formed at 285°K.

The decrease in quantum efficiency at high concentrations can be tentatively explained if the type of V trap changes. If the traps become predominantly of the V_3 type (two positive ion vacancies) rather than the V_2 type or V_4 type (respectively two positive ion vacancies with one positive hole, and two positive ion vacancies with one negative ion vacancy), σ_p will decrease and σ_v/σ_p increase, as required. At the same time, because of the electrical equilibrium of the crystal, there will necessarily be a residual number of empty negative ion vacancies. Thus v will exceed N.

The variation of the quantum efficiency of bleaching with temperature for the coloration density at which fig. 5 was obtained may be written in terms of (3)

$$\frac{fI_0(1-10^{-D})}{K\,dD/dt}-1=\frac{H}{G}\exp\frac{E}{kT}. \qquad (4)$$

The straight line drawn in fig. 5 has the values $H/G=0{\cdot}028$ an $E\mathrm{d}=0{\cdot}043$ ev. The experimental plot is inaccurate owing to the small value of the left-hand-side of (4) at room temperature. More accurate values of H/G and E are being obtained by extending the temperature range down to that of liquid helium. For additively coloured crystals Mott and Gurney (1948, p. 136) derive $H/G=0{\cdot}0033$ and $E=0{\cdot}075$ ev. The values found from fig. 5 allow one to say that a similar activation energy exists in x-ray coloured crystals. This is significant in showing that the F centre behaves normally and is unassociated with positive holes. This is in agreement with the observation of photoconductivity in x-ray coloured crystals (Oberly 1951).

The creation of F centres by x-rays will now be considered, and equations for the process suggested. From these it appears that the trapping of electrons at frozen-in ion vacancies is inadequate to explain the experimental result, and the equations are extended to include trapping at incipient vacancies.

The primary action of the x-rays is to liberate electrons and positive holes within the crystal, and these may annihilate each other or become trapped at imperfections. If e is the concentration of free electrons, X the concentration of trapping sites (including those occupied by electrons), n the concentration of occupied traps, σ_X the capture cross section of an empty trap, and u an electron velocity parameter; the rate at which electrons are trapped is $u\sigma_X e(X-n)$. Similarly the rate of annihilation of electrons by positive holes is $u\sigma_P ep$. Suppose also that thermal spikes caused by the X irradiation release electrons from traps at a rate an per unit volume, and that x-rays liberate electrons at a rate JI_x, where I_x is the x-ray intensity. If there is no other process the rate of creation of free electrons is

$$de/dt=JI_x-ue(\sigma_p p+\sigma_X(X-n))+an \qquad (5)$$

and the rate of creation of trapped electrons is

$$dn/dt=u\sigma_X e(X-n)-an. \qquad (6)$$

Now a calculation based on Pohl's figures (Mott and Gurney 1948, p. 127) for the range of photoelectrons in additively coloured crystals, and on Harten's data on electrical conductivity during X irradiation, suggests that the concentration of free electrons is $10^3\,\mathrm{cm}^{-3}$ when the F centre concentration is $10^{16}\,\mathrm{cm}^{-3}$. Thus, if the traps represent negative ion vacancies, and the trapped electrons F centres, one may take $de/dt\approx 0$, so that (6) becomes

$$\frac{dn}{dt}=\frac{\sigma_X(X-n)(JI_x+an)}{\sigma_p p+\sigma_X(X-n)}-an \qquad (7)$$

(7) should give the experimental eqn. (1). This is possible only if there is an improbable relation between the constants, or if X is constant and much greater than the maximum value of n, i.e. $X \gg 10^{17}$ cm $^{-3}$. This is unacceptable on two grounds. Firstly, the concentration of negative ion vacancies is expected to be not greater than 10^{17} cm^{-3} (Seitz 1954, p. 14), and secondly, Sakaguchi and Suita's experiments show that the concentration is not constant, but increases as F centres are created.

It is thus necessary to assume that the primary trapping site is not a negative ion vacancy, but that F centres can be generated from the occupied traps. Since the volume expansion appears to proceed by a similar curve to F centre creation, it appears reasonable to assume that the trap is an incipient negative ion vacancy which, when it has captured an electron, may expand to form an F centre. In this case one may suppose that, in addition to the processes already listed, incipient centres become F centres at a rate bn, that electrons are liberated from F centres by thermal spikes at a rate cN, and that F centres revert to incipient centres at a rate gN. It is assumed that the positive holes are immobile, and the recombination of holes with incipient centres and F centres is therefore ignored. Equations (5) and (6) thus become:

$$de/dt = JI_x + an + cN - ue(\sigma_p p + \sigma_X(X-n)). \quad . \quad . \quad . \quad . \quad (8)$$

and

$$dn/dt = u\sigma_X e(X-n) - an - bn + gN. \quad . \quad . \quad . \quad . \quad . \quad (9)$$

In addition the rate of formation of F centres is

$$dN/dt = bn - cN - gN. \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

These equations only have a solution of type (1) if the concentration of free electrons, e, is constant, an assumption which Harten's data indicates to be approximately true; if the concentration of occupied traps, n, is also constant, and if F centres once formed do not revert to incipient centres. In this case B is the rate of liberation of electrons from F centres, and A is a function of σ_p, σ_X, u, a, b, J, I_x, and X. A may be written less cumbrously in quantities other than I_x and (1) then has the form

$$\frac{dN}{dt} = \frac{bX}{1+(a+b)\sigma_p/c\sigma_X} - cN. \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Since it is experimentally observed that as the temperature rises the rate of creation of F centres approaches the rate of liberation of electrons, it may be assumed that incipient centres become F centres, rather than being ionized, i.e. $b > a$. Thus the temperature dependence of A will be a function of that of b, c, and σ_p/σ_X, and its small value is not unreasonable. The presence of X in the numerator of (11) accords with Gordon and Nowick's observation that plastic deformation increases both the rate of coloration and the saturation coloration. If X refers to the concentration of incipient negative ion vacancies at jogs on dislocations, it is necessary to assume evaporation of incipient positive and negative ion vacancies

so that X will not be reduced as F centres are formed, otherwise X would have to be implausibly large to provide an F centre concentration of $7 \times 10^{16}\,\mathrm{cm}^{-3}$. If an incipient negative ion vacancy evaporates and forms an F centre after trapping an electron, it seems likely that an incipient positive ion vacancy may evaporate after trapping a positive hole, thus leaving an incipient negative ion vacancy at the jog again.

From the conditions required for a linear solution of (8), (9) and (10) it is expected that n will be much smaller than X, and the experimental evidence does not conflict with this. At the end of X irradiation the n incipient centres might become F centres, though perhaps at a rate slower than in the presence of X irradiation, leading to a continued increase in F centre concentration. Attempts to measure such an increase at room temperature give a value of $10^{13} \pm 3 \times 10^{13}\,\mathrm{centres\,cm}^{-3}$, so that n may very well be much less than $10^{13}\,\mathrm{cm}^{-3}$.

All the results discussed are to a good approximation reproducible in successive X irradiations, but in a run of 25 cycles of X irradiation and optical bleaching at room temperature it was observed that the initial rate of growth of F centres increased by about 80%. This ageing effect has not been further studied but it is affected by a comparatively small amount of annealing; an increase of 20°C for less than one hour producing a decrease of 10% in the subsequent growth rate. It appears that temperature control to within one or two degrees is necessary for quantitative study.

§ 6. Summary

Both sets of experiments provide evidence for the mobilization of single negative ion vacancies by X irradiation. It is thought that creation of vacancies occurs by the evaporation, from jogs in dislocations, of incipient ion vacancies which have trapped electrons. The vacancies at which the F centres are formed are only stable when occupied by electrons and disappear when these are optically released to the conduction band. The saturation concentration of F centres is governed by the rate at which electrons are released from F centres by the X irradiation, as well as by the concentration of trapping sites. The constants of the thermal process by which electrons are raised from the excited state to the conduction band are considered to show that the environment of the F centre is the same as in additively coloured crystals, and that the positive hole does not affect the behaviour of the F centre.

Acknowledgments

The author is grateful to Dr. D. A. Jones of Aberdeen University for the gift of the potassium chloride crystal, to Dr. H. Mykura and Dr. G. A. P. Wyllie for helpful discussion during the course of work, and to Dr. E. E. Schneider of the University of Durham for valuable criticism of the manuscript.

References

Duerig, W. H., and Mador, I. L., 1952, *Rev. sci. Instrum.*, **23**, 421.
Gordon, R. B., and Norwick, A. S., 1956, *Phys. Rev.*, **101**, 977.
Harten, H. U., 1949, *Z. Phys.*, **126**, 619.
Herman, R., and Wallis, R. F., 1955, *Phys. Rev.*, **99**, 435.
Hesketh, R. V., 1958, *J. sci. Instrum.* (in the press).
Mott, N. F., and Gurney, R. W., 1948, *Electronic processes in ionic crystals* (Oxford : University Press).
Oberly, J. J., 1951, *Phys. Rev.*, **84**, 1257.
Seitz, F., 1954, *Revs. mod. Phys.*, **26**, 7.
Theimer, O., 1958, *Phys. Rev.*, **109**, 1095.

A Study of the Palaeomagnetism of the Bushveld Gabbro†

By D. I. Gough and C. B. van Niekerk
National Physical Research Laboratory,
Council for Scientific and Industrial Research, Pretoria

[Received November 7, 1958]

Abstract

Data are presented on the natural remanent magnetization at five sites of the Main Gabbro zone of the Bushveld Igneous Complex. It is found that the directions of N.R.M. at the five sites would agree more closely if the rock at each site were rotated about the strike of the pseudo-stratification so as to make its dip zero. This indicates that the gabbro became magnetized before the crust subsided to form the basin in which the Complex now lies. The same results show that the N.R.M. of the gabbro has been stable since before the crustal subsidence. Stability of the N.R.M. is further demonstrated by the highly consistent directions of magnetization making a large angle with the present field. After correction for dip, and on the assumption that the gabbro was magnetized in the field of a geocentric dipole, the mean north magnetic pole from the five sites lies at 23° N. 36° E. The age of the Complex is close to $2{\cdot}0 \times 10^9$ years. The palaeomagnetic results show that the crust supported the load due to the gabbro intrusion without subsidence for at least the time taken by this intrusion to cool to the Curie temperature of its ferrimagnetic constituents. A lower limit to this time is estimated, on reasonable assumptions, to be of the order of 10^5 years, so that the South African crust 2×10^9 years ago appears to have had very considerable strength.

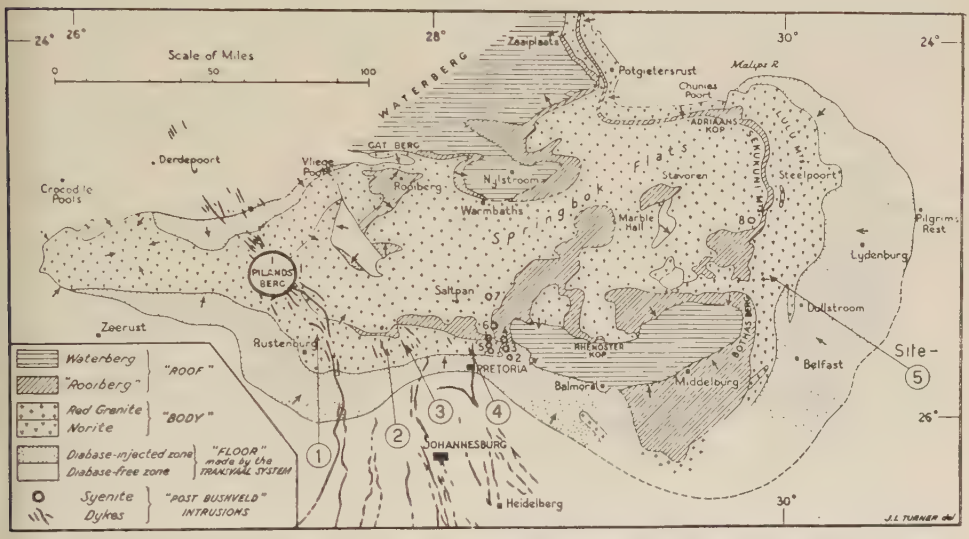
§ 1. Introduction

The Bushveld Igneous Complex is one of the largest systems of intrusive rocks known. Its main features are described by du Toit (1954). Figure 1 shows its general structure, as given by du Toit. A basin, some 200 miles long from east to west and about 90 miles from north to south, is filled principally with gabbro (often described as norite, as in fig. 1) and red granite, the gabbro being exposed round the flanks of the basin with the granite in the centre. The total area of gabbroic rock is nearly 20 000 sq. miles. The metamorphosed sediments of the Transvaal System which surround and underlie the gabbro, show systematic dips towards the centre of the Complex. In its lower levels the gabbro itself is differentiated into banded mineral assemblages which are inclined like the underlying sediments. Higher in the intrusion, pseudo-stratification in the undifferentiated gabbro shows similar, though less obvious, evidence of dip towards the centre of the basin. Such data have led to the conclusion that the intrusion caused an extensive subsidence of the crust at some time after crystallization was well advanced.

† Communicated by Professor P. M. S. Blackett, F.R.S.

The age of the granite of the Complex has recently been firmly established. Schreiner (1958) used rubidium–strontium decay measured in total rock and separated mineral samples. He gives the age 1920 million years for the Bushveld granite, using half-life $5{\cdot}0 \times 10^{10}$ years, with 99% confidence limits ± 130 million years. Nicolaysen *et al.* (in the press) used rubidium–strontium and uranium–lead decay in various minerals from the Bushveld granite and give a mean age of 1950 ± 150 million years. A specimen of biotite-bearing pyroxenitic gabbro from the Merensky Reef in the differentiated zone of the gabbro was found to be 2050 ± 50 million years old (Nicolaysen *et al.*). The Bushveld Complex is therefore one of the best dated of ancient rock systems. The age of the gabbro must be close to $2{\cdot}0 \times 10^9$ years. The known age, together with the relative accuracy with which strike and dip could be defined round the basin, led to the palaeomagnetic study here reported.

Fig. 1



Simplified geological map of the Bushveld Igneous Complex, after A. L. du Toit. The five sites sampled in the gabbro are indicated.

§ 2. Sampling and Measurement

Samples were cores of diameter one-half inch drilled from the rock at the site by means of a portable field drill fitted with a diamond coring crown. Before removal each core was oriented in one of several ways. Most of the surface cores were oriented with reference to the sun's azimuth, some with reference to survey beacons and a few from a magnetic compass, which was not appreciably deflected by the gabbro. The underground cores from chrome mines were oriented relative to the mine survey networks. From one to four specimens one-half inch long were cut from each core in the laboratory, the number being two in nearly all cases.

The natural remanent magnetization (N.R.M.) of each specimen was measured in direction and intensity by means of an electromagnetic 'spinner' type magnetometer. The specimens from the Bon Accord quarries (Site 4) were measured with the magnetometer used for the study of the Pilanesberg dykes (Gough 1956). All other specimens were measured with a new and much improved spinner magnetometer which will be described elsewhere.

Mean directions were found for each core from the specimens (normally two) cut from it, and the unit in the statistical data and in the stereograms of fig. 2 is the mean direction for a core.

§ 3. Field Work

It has been found that, in general, rock samples from the surface in South Africa are of little or no use for palaeomagnetic studies (Gough 1956, Graham and Hales 1957). Attention was therefore confined to exposures of fresh rock in excavations. A set of cores from a quarry in the Main Gabbro zone at Bon Accord, north of Pretoria (see fig. 1, Site 4) gave north-seeking poles directed near the downward vertical, as shown in fig. 2. This result immediately raised the question whether the directions of N.R.M. should be corrected for the dip of the gabbro, that is, whether the N.R.M. observed had been acquired with the rock in its present attitude or in some other. The obvious way to answer this question was to sample the formation elsewhere where its strike and dip were as widely different as possible from those at Bon Accord. Cores were therefore collected from chrome mines in the north-east and north-west flanks of the basin. No undifferentiated gabbro or norite was exposed in these mines, and the pyroxenite cores were found to be feebly magnetized with scattered directions. A set of cores from a dam site (Site 5) near Roossenekal, in the eastern flank of the Main Gabbro zone showed close grouping of directions of N.R.M. Thereafter sampling was confined to exposures of the Main Gabbro zone. Five sites were sampled, including the two already mentioned. They are marked in fig. 1. Site 1 is a group of three quarries near Rustenburg, and Sites 2 and 3 are quarries near Brits, all in the Main Gabbro. Careful search failed to discover any suitable exposures of gabbro in the west, north-west or north-east flanks of the basin.

Table 1 gives the positions of the five sites, and the strike and dip of the pseudo-stratification at each. The determinations of dip and strike presented problems because the rock at this horizon is homogeneous and shows no bands of mineral assemblages such as produce visible layering lower in the gabbro. Lenticular xenoliths of quartzite, which are oriented parallel to the pseudo-stratification, were well exposed only in one quarry. The rock develops natural fracture planes on weathering which follow the pseudo-stratification, and where such planes were freshly exposed in quarry excavations they were used for determination of dip and strike. Similar fracture surfaces exposed as outcrops at the surface were found

to be more or less rounded by weathering, and gave systematically low dips. They were therefore used only for supplementary data on the strike. Blasting in the quarries produces splitting along parallel planes which tend to follow the pseudo-stratification. Such blasting fracture surfaces were used in the same way as natural fracture planes for measurement of strike and dip.

Table 1. Positions of the Sites and Strike and Dip at each

Site No.	Description	Latitude (°S.)	Longitude (°E.)	Strike (°E. of N.)	Dip (°)
1	Three quarries near Bospoortdam, Rustenburg district	25·54	27·35	152 ± 7	11 ± 1 E.NE.
2	Quarry at Rooikoppies, Brits district	25·62	27·74	91 ± 1	$9{\cdot}9 \pm 0{\cdot}7$ N.
3	Quarry at Elandsfontein, Brits district	25·62	27·84	107 ± 4	12 ± 1 N.
4	Two quarries at Bon Accord, Pretoria district	25·62	28·21	90 ± 5	30·5 N.
5	Dam site near Roossenekal	25·27	29·95	0	$13{\cdot}7 \pm 0{\cdot}7$ W.

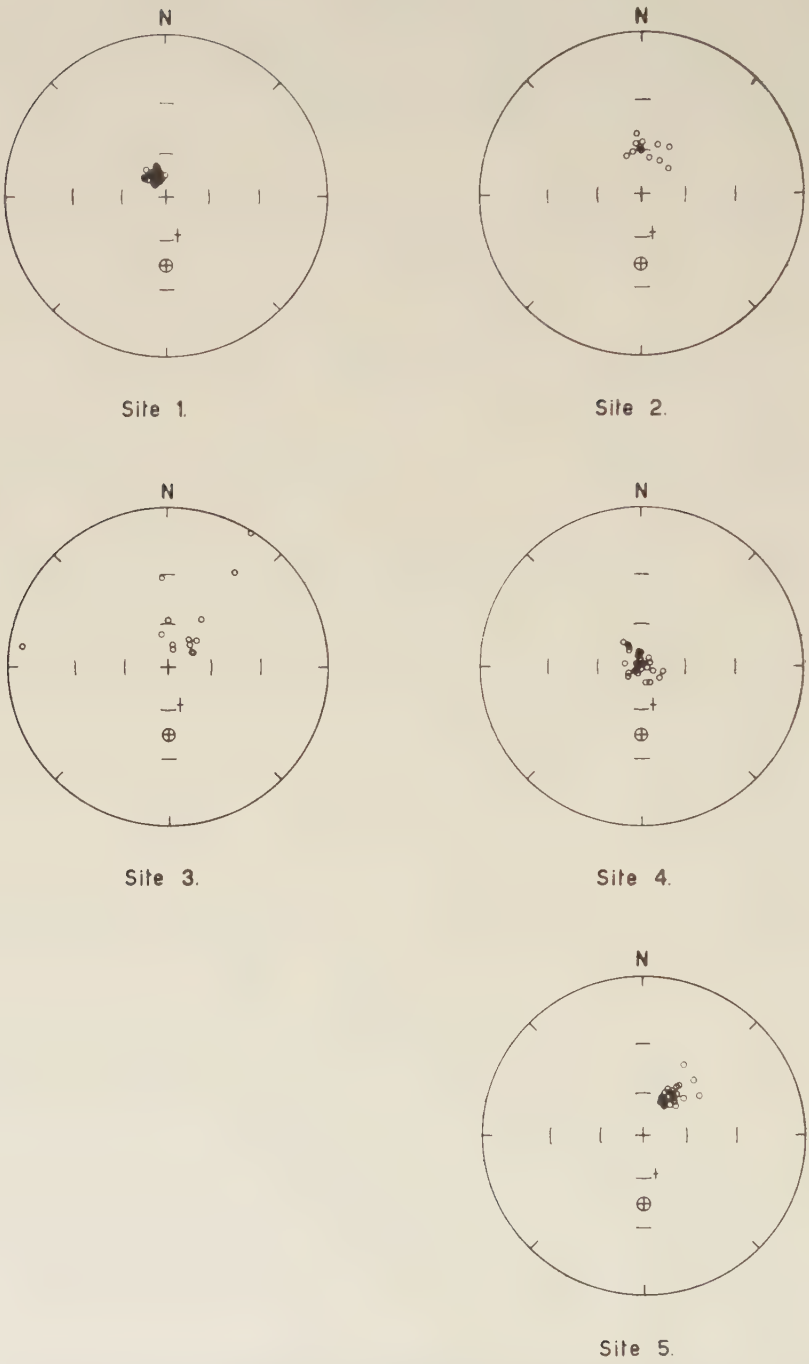
Table 1 shows that except at Site 4, dips were all small. Individual measurements of strike and dip showed variations which are reflected in the standard errors of the mean values given in table 1.

§ 4. Natural Remanent Magnetization of the Bushveld Gabbro

The directions of N.R.M. are represented stereographically in fig. 2. All the cores had north-seeking poles in the lower hemisphere of projection. It will be seen that the directions found at Sites 1, 4 and 5 were very closely grouped. Sites 2 and 3 gave less closely grouped directions. At Site 3 four cores whose N.R.M. vectors fell outside the main concentration were omitted from the statistical computation. These cores were nearer than the others to a weathered surface. At the other four sites all cores measured were included in the analysis, except two whose orientations were lost after drilling.

At Site 1 one core was drilled from gabbro into a xenolith of quartzite, and one gabbro specimen was cut from it, two quartzite, and one composite. The directions of N.R.M. were closely grouped and the mean

Fig. 2



Directions of natural remanent magnetization at five sites in the Bushveld Gabbro. Lower hemisphere stereographic projection.

O North-seeking poles of rock cores.
⊕ Axial dipole field (south-seeking).
† Present geomagnetic field (south-seeking).

of the four was used as one core direction in the analysis. The agreement indicates that the gabbro raised the quartzite to above the Curie temperature of its ferrimagnetic constituent. The two quartzite specimens had intensities of N.R.M. 0·13 and 0·043 of that of the gabbro specimen.

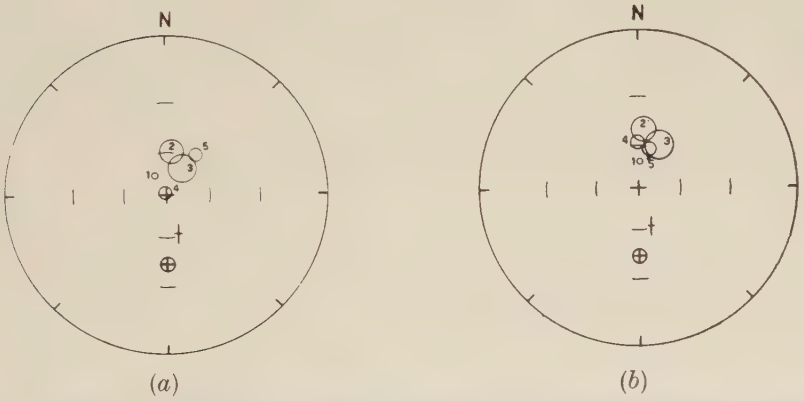
The results of the measurements of N.R.M. are given in table 2, which gives the mean direction of N.R.M. for each site, the angular radius of the cone of 99% confidence for the mean, and the parameter K, using Fisher's statistics (Fisher 1953). The confidence cones are stereographically

Table 2. Statistics of N.R.M. at Five Sites

Site	N	N'	Mean direction		α 99%	K
			Inclination	Azimuth (E. of N.)		
1	29	0	+ 73°·8	330°·0	2°·0	268·2
2	12	0	+ 59°·3	6°·9	7°·8	51·5
3	10	4	+ 68°·0	30°·2	9°·6	43·5
4	29	0	+ 87°·7	323°·9	4°·4	60·7
5	23	0	+ 55°·3	36°·8	4°·4	78·5

N = number of cores included in analysis.
N' = number of cores excluded from analysis.
α = semi-angle of cone of 99% confidence for mean direction.
$K = N - 1/N - R$ where R is the sum of N unit vectors.

Fig. 3

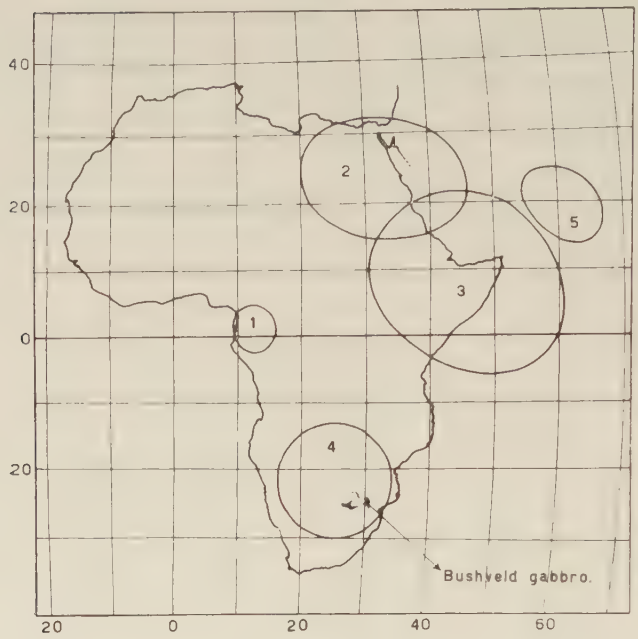


(*a*) (*b*)

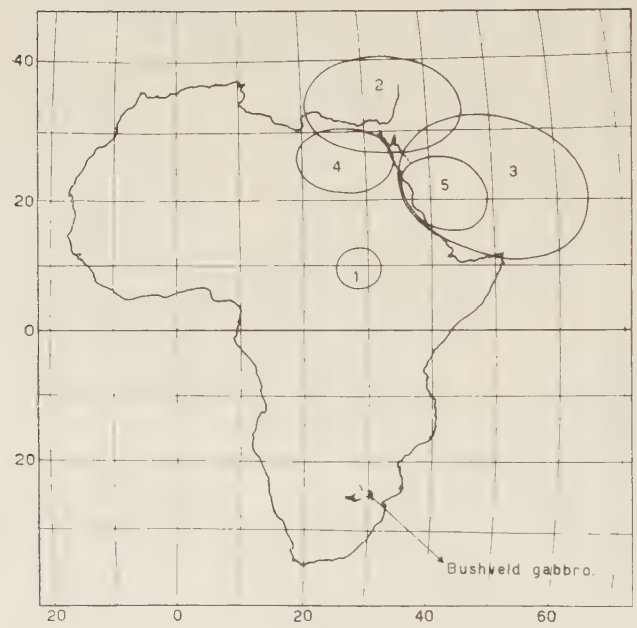
Circles of 99% confidence for mean directions of magnetization at five sites in the Bushveld Gabbro. (*a*) with the rock at each site in its present attitude ; (*b*) after correction for dip of pseudo-stratification.

projected in fig. 3 (*a*). In fig. 3 (*b*) the confidence circle for each site has been moved to the position it would occupy if the gabbro at the site were rotated about the strike as axis so as to reduce the dip to zero. It will be seen that the mean directions for the five sites are more closely grouped after correction for dip than before. There are significant residual

Fig. 4



(*a*)



(*b*)

Ovals of 99% confidence for the pole positions given by five sites in the Bushveld Gabbro, on the assumption of a geocentric dipole field. (*a*) with the rock in its present attitude at each site ; (*b*) after correction for dip.

differences in fig. 3 (*b*), which are no doubt attributable partly to errors in the dip and strike corrections and partly to secular variation of the geomagnetic field between the dates of passage through the Curie temperature at the various sites. Sites 1 and 5 are about 160 miles apart, and it would be reasonable to suppose that the gabbro at one of them might have become magnetized hundreds of years later than that at the other.

Figure 4 (*a*) shows the oval of 99% confidence for the position of the North magnetic pole given by the data from each site, on the assumption that the observed N.R.M. was produced by the field of a geocentric dipole. Figure 4 (*b*) gives the corresponding North magnetic pole positions after correction for dip. It is clear from figs. 3 and 4 that the consistency of the data improves markedly when the pseudo-stratification is made horizontal at each site. This amounts to Graham's stability test on a large scale (Graham 1949). The simplest explanation is that the gabbro acquired its N.R.M. before the development of the crustal subsidence and that the N.R.M. has remained stable since the subsidence took place. It is possible that subsidence did not take place until after the passage of an appreciable fraction of the total time since the intrusion of the gabbro. Apart from the evidence of stability afforded by the subsidence, there is a strong argument for stability in the consistent and closely grouped N.R.M. vectors making an angle of about 120° with the present geomagnetic field. It is reasonable to suppose that the observed N.R.M. is the original thermo-remanent magnetization acquired when the gabbro cooled.

The mean positions of the North magnetic pole, derived from the five sites after correction for dip on the assumption of a geocentric dipole field, are given in table 3. The mean of the five mean pole positions with equal weights gives a North magnetic pole at latitude 23° N., longitude 36° E. There is relatively little difference between this and the pole position at the time of the intrusion of the Pilanesberg dykes.

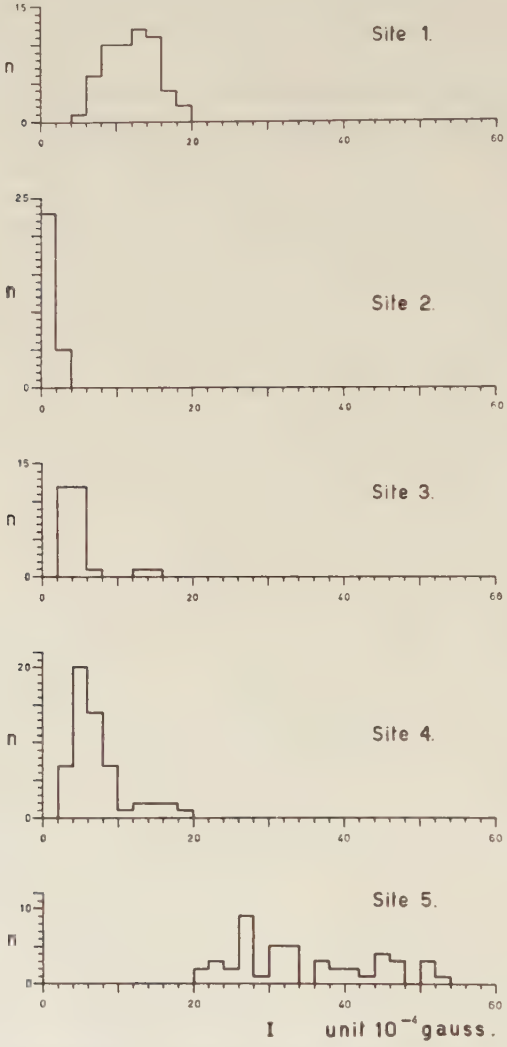
Table 3. Mean Position for North Magnetic Pole

Site	Pole position	
	Latitude	Longitude
1	+10	+28
2	+34	+33
3	+23	+50
4	+26	+26
5	+20	+42
All	+23	+36

The distribution of intensity of N.R.M. at each site is shown in fig. 5. The unit here is the single specimen so that the total number of data for each site is about double that for the directions.

The palaeomagnetic data afford evidence that the subsidence of the crust, which formed the present basin, did not develop until the rock had cooled to below the Curie temperature of its ferrimagnetic constituent.

Fig. 5



Distribution of intensities of natural remanent magnetization at five sites in the Bushveld Gabbro.

The Curie temperature for pure magnetite is 585°c and for titanomagnetites is less. It is therefore a conclusion from the work here reported that the gabbro cooled to 600°c or less before the crust subsided. It is amusing to attempt an estimate of the cooling time. The Main Gabbro forms a homogeneous body of rock which may reasonably be supposed

to have been intruded rapidly enough so that its whole thickness was near the magma temperature at one time. In order to derive a minimum time for cooling of the Main Gabbro it will be assumed that the highly differentiated Critical Zone of the gabbro was already in place below it. Above the Main Gabbro was cap-rock whose position is now occupied by granite. The inclinations of the Transvaal System floor, of the Critical Zone, of the gabbro and of the granite are similar, and it is reasonable to suppose that the Main Gabbro formed a parallel sided slab after intrusion. If the cap-rock was much thicker than the gabbro slab, a lower limit for the cooling time would be given by taking the cap-rock above and the Critical Zone below to have been at zero temperature when the gabbro was intruded. There is in fact evidence that the Critical Zone was still at a high temperature. Du Toit (1954, p. 184) gives 15000 feet as a representative thickness for the Main Gabbro zone. The five sites described in this paper are near the centre of the Main Gabbro. Assuming that the Main Gabbro was initially at 1000°C and had great thicknesses of rock at ordinary temperatures above and below it, an estimate can be made of the time required for it to cool to 600°C. Jaeger has calculated curves for this case, with assumed values of the thermal constants, which give

$$\text{time in years} = 0{\cdot}016D^2$$

for the temperature near the centre of such a slab to fall from 1000°C to 600°C, where D is the thickness in metres. At a point $D/20$ from the contact the corresponding time is $0{\cdot}010D^2$, so that neither convection nor the position within the slab affects the order of magnitude. With $D = 4700$ m, the time is $3{\cdot}5 \times 10^5$ years. The simplifying assumptions which do not minimize the time are (1) that the cap-rock was thick compared with the intrusion, (2) that the intrusion extended beyond the sites at least two or three times its thickness. Unless these assumptions are greatly in error it seems reasonable to conclude that the gabbro at the sampling sites took longer than a time of the order of 10^5 years to cool from 1000°C to 600°C. The crust apparently supported the gabbro for a time greater than this without subsiding into the present basin. This suggests that the crust in South Africa 2×10^9 years ago had very considerable strength. There is of course no evidence from the present work to show whether a block including the whole of the Bushveld Complex moved so as to reduce the stresses.

Acknowledgments

We wish to thank Dr. E. J. Marais for his support and Dr. A. L. Hales for helpful discussions. We have been assisted by the willing cooperation of the staffs of various quarries and mines, who are too many for us to enumerate here. The work here reported was done as an official project of the National Physical Research Laboratory, Council for Scientific and Industrial Research, and we acknowledge the permission of the C.S.I.R. to publish this paper.

We are indebted to Messrs. Oliver and Boyd and to the executors of the late Dr. A. L. du Toit for permission to reproduce as fig. 1 a geological map from Du Toit's *Geology of South Africa*.

References

du Toit, A. L., 1954, *Geology of South Africa*, 3rd edn.
Fisher, Sir Ronald, 1953, *Proc. roy. Soc.* A, **217,** 295.
Gough, D. I., 1956, *Mon. Not. R. astr. Soc., geophys. suppl.*, **7,** 196.
Graham, J. W., 1949, *J. geophys. Res.*, **54,** 131.
Graham, K. W., and Hales, A. L., 1957, *Advanc. Phys.*, **6,** 22, 149.
Nicolaysen, L. O., de Villiers, J. W. L., Burger, A. J., and Strelow, F. W. E., *Trans. geol. Soc. S. Afr.* (in the press).
Schreiner, G. D. L., 1958, *Proc. roy. Soc.* A, **245,** 112.

CORRESPONDENCE

Further Observations of Stroh-cracks in Magnesium Oxide Single Crystals

By R. J. Stokes, T. L. Johnston and C. H. Li
Honeywell Research Center, Hopkins, Minnesota, U.S.A.

[Received August 21, 1958]

In a previous paper (Stokes *et al.* 1958) we have reported the existence of 'Stroh-cracks' (defined as cracks lying in (110) planes perpendicular to the active glide planes) in magnesium oxide single crystals. These cracks were found in regions of locally severe non-homogeneous deformation, formed as a result of buckling under simple compression. As was pointed out, this was rather a special mode of deformation. Since that time we have become more familiar with this material and now appreciate that Stroh-cracks are indeed a very frequent occurrence and can be observed even after simple bending. In this note we present some recent interesting observations.

Single crystals with cross section $\frac{1}{4}$ in. $\times \frac{1}{8}$ in. and 1 in. in length were chemically polished by boiling for one minute in fresh 85% ortho phosphoric acid. After being washed in distilled water and dried with methyl alcohol, they were transferred to a testing machine and deformed under a three-point bending load. Crystals having this treatment were generally very ductile and showed little work hardening. Thus, the stress-deflection curves were almost flat in the plastic region, although periodically there were small yield drops giving a form of 'jerky flow'. The stress distribution along the crystal beam was such that it reached its maximum value in the centre where the load was being applied, hence all of the plastic deformation was confined to this region. In other words, we again have a condition where the deformation can be both severe and non-homogeneous. This time, however, the stress field changes from compression to tension across the neutral axis.

As was to be expected, crystal specimens which had been deformed in this way contained Stroh-cracks. Furthermore, as previously reported, these Stroh-cracks generally extended right through the crystal in the form of tiny slits. These slits lay along the [100] direction parallel to the axis of bending. Figure 1 (*a*)† is a photomacrograph taken on a specimen deformed all the way to fracture. Although one slit towards the neutral

† All figures are plates.

axis stands out very prominantly in the photograph, several others can be seen closer to the compression surface itself. Figure 1 (*b*), on the other hand, was taken on a specimen which had not yet fractured. In this photomacrograph two slits can be distinguished in that half of the crystal which was *under tension*. The existence of a stable Stroh-crack under tension in a material such as magnesium oxide was rather surprising.

The photomicrographs reproduced in fig. 2 were obtained on the same specimen as fig. 1 (*b*), they show the situation and orientation of Stroh-cracks in the tension region before and after failure. Figure 2 (*a*) was taken when the specimen had a nominal surface strain of 7·5% at the centre of the beam. This strain was calculated on the assumption of an elastic beam, the actual plastic strain was much greater (probably 15%). In fig. 2 (*b*), taken after failure, the fracture surface can be seen to have passed through the larger of the two Stroh-cracks. When the main fracture path passes through the tiny slits, narrow steps are formed in the fracture surface whose orientation lies parallel to a (110) slip plane rather than the more characteristic (100) cleavage plane. Closer examination of the fracture surfaces of a number of ductile specimens frequently revealed such a step in the tension region. In fact, the crystal of fig. 1 (*a*) had one; its profile is indicated by the arrow.

The situation and orientation of the Stroh-cracks in fig. 2 (*a*) and fig. 3 provide further confirmation for the mechanism put forward in the original paper to explain their development. It can be seen that the cracks originate in an area where slip planes intersect. However, not every such area in fig. 3 for example develops a crack, so there must be some special feature which dictates where they form. This special feature again is the kink band. Figure 2 (*a*) shows the two Stroh-crack components lying along a boundary of the kink band ; they have formed where it is intersected by an active slip plane. Once having formed, their orientation switches to the cleavage plane over which there is a tension stress, thus accounting for their S-shape. The larger crack also contained a ' secondary crack ' segment (Stokes *et al.* 1958) going across the kink band at one end. It should be pointed out that the cleavage plane component at the other end of the larger crack of fig. 2 (*a*) also went right through the crystal as a part of the stable slit.

Details of the stages involved in the nucleation and growth of Stroh-cracks will be discussed in a future paper.

Acknowledgments

This work is supported by the Office of Naval Research. The authors wish to thank Mr. K. H. Olsen for his assistance.

Reference

Stokes, R. J., Johnston, T. L., and Li, C. H., 1958, *Phil. Mag.*, **3**, 718.

The Ejection of Atoms from Gold Crystals During Proton Irradiation

By M. W. Thompson
Metallurgy Division, A.E.R.E., Harwell

[Received September 16, 1958]

Certain experiments on irradiation damage imply that interstitial atoms are produced at a considerable distance from vacant lattice sites (Kinchin and Pease 1955, Thompson *et al.* 1957). Silsbee (1957) has suggested that under the right conditions of atomic size and lattice parameter, the momentum transfer in successive atomic collisions might be aligned along the close-packed directions of the crystal. Leibfried (1958) has shown that this effect might permit the transmission of energy pulses over distances of the order 100 Å. This letter is a preliminary report on an investigation into the directions in which atoms leave the surface of a gold crystal under irradiation.

In fig. 1 the general arrangement of the apparatus is shown. A beam of protons from a Van de Graaff accelerator passed through the atmosphere into a thin gold foil. This foil was prepared by rolling to 0·002 in. from a $\frac{1}{8}$ in. rod followed by annealing at 850°C for 3 hours. A strongly preferred orientation of grains was thus achieved, the [100] planes being in the surface and the (001) direction in the rolling direction. This is the well-known ' cubic texture ' (see for example Barrett 1943).

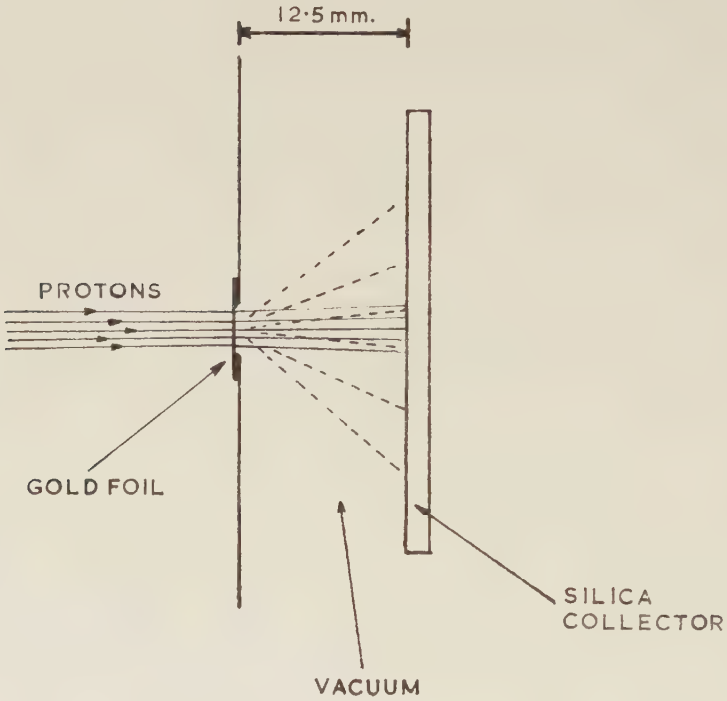
Under irradiation the foil was cooled with an air blast so that its maximum temperature was below 200°C, eliminating any possibility of evaporation. Behind the foil was an evacuated chamber at 10^{-5} mm pressure containing a silica plate coated with about 100 Å of aluminium and cooled to −150°C upon which the gold atoms were collected. The purpose of the aluminium coating was to enhance the sticking of gold atoms and to provide a collector at the same potential as the target. Before irradiation the inner surface of the foil was cleaned by admitting 0·1 mm of argon to the vacuum chamber and making the foil the cathode of a gas discharge. During this process the collector was screened to prevent sputtered atoms from condensing. To ensure that the surface remained clean during irradiation a baffle system cooled to −150°C surrounded the target region.

The proton energy was adjusted so that the full beam current of 3 microamp just penetrated the foil, the mean energy of the emerging protons was then about 0·3 Mev. A total dose of $3{\cdot}4 \times 10^{16}$ protons was given to the foil over an area of 0·1 cm^2 after which the collector plate was removed and exposed to a neutron flux of 10^{12} neutron/cm^2/sec for $2\frac{1}{2}$ days.

After allowing 4 days for the short-period activity of the aluminium and silica to decay, the gold deposit which had a $2\frac{1}{2}$-day half-life could be detected by autoradiography or with a Geiger counter.

In fig. 2, Pl. 31 an autoradiograph is reproduced which was made on Ilford 'G' x-ray film exposed for 32 hours. This shows a cross-shaped deposit of gold against a background of impurity particles with long half-lives. The total number of gold atoms collected was determined by

Fig. 1



measuring the activity with a Geiger counter and found to be approximately 10^{13}. The arms of the cross and their orientation relative to the rolling direction of the foil correspond to preferential ejection in the crystal directions (110), (1$\bar{1}$0), (101) and (10$\bar{1}$). Since gold has a f.c.c. structure these are also the close-packed directions.

An interpretation of this experiment is that the momentum transferred in the initial proton–atom collision is resolved into pulses travelling along the close-packed lines of atoms. These pulses would be focused into the lines by the effect proposed by Silsbee. Provided that a particular pulse is not too far attenuated by the time it reaches the surface the last atom would be ejected.

Acknowledgments

I am indebted to Dr. M. Adams for his advice on the preparation of textured foils, Dr. E. B. Paul and members of the Van de Graaff group for their cooperation and to Miss J. Stubbs and Mr. B. Farmery for their technical assistance.

References

Barrett, C. S., 1943, *The Structure of Metals* (New York : McGraw-Hill).
Kinchin, G. H., and Pease, R. S., 1955, *J. nuclear Energy*, **1**, 200.
Leibfried, G., 1958 (to be published).
Silsbee, R. H., 1957, *J. appl. Phys.*, **28**, 1246.
Thompson, D. O., Blewitt, T. H., and Holmes, D. K., 1957, *J. appl. Phys.*, **28**, 742.

REVIEWS OF BOOKS

A Handbook of Structures and Lattice Spacings of Metals and Alloys. By W. B. Pearson. (London: Pergamon Press, 1958.) [Pp. 1044.] Price £13 2*s*. 6*d*.

The author remarks in his preface that his book has been written to assist the physicist and metallurgist by providing a brief reference book which contains information about the structures and lattice spacing of all binary and ternary alloys which have been examined. It is a pity that it has not been possible to publish this at a price which will enable many physicists or metallurgists to possess it.

The book has beeen written in two parts which might well have been published as separate volumes. Part I contains a brief account of the Debye Scherrer method of measuring lattice spacings together with some short chapters on the relationship of lattice spacing measurements to such topics as the determination of equilibrium diagrams, the theory of alloy formation. and the magnetic properties of metals and alloys. Part II, which is the main body of the work, sets forth in an easily accessible way, firstly structural details of all known binary and ternary phases; and subsequently the results and sources of all x-ray work on metals and alloys published before 1956, together with important results obtained by other methods and some more recent work. A final chapter gives details of work on carbides, hydrides, nitrides and oxides. This part of the work is notable for the amount of work which is covered, and will prove of value to all who are interested in the properties of metals and their alloys.

P. J. B.

Nuclear Quadrupole Resonance Spectroscopy—Solid State Physics, Supplement 1. By T. P. Das and E. L. Hahn. (London : Academic Books Ltd.) [Pp. ix+223.] 56*s*.

Nuclear quadrupole resonance spectroscopy is now employed more and more as an analytical tool in studies of the solid state, and structural chemistry. This volume is issued as the first supplement to the Academic Press *Solid State Physics* series and presents systematically the theory, experimental techniques, and some important experimental examples of nuclear quadrupole studies. The text emphasizes especially the pure nuclear quadrupole interactions with crystalline and molecular electric fields. Nuclear magnetic moment interactions are considered only in the minor role of perturbations upon the electric quadrupole energy levels, owing to the fairly extensive literature concerning magnetic effects which is already available.

The first part of the book discusses concisely the frequencies and intensities of Pure Quadrupole Spectra, Zeeman splitting in axially symmetric field gradients and then in the general case of crystals not possessing this axial symmetry. ' Real life ' crystalline conditions are next considered, with the possibilities of static dipole–dipole and indirect spin–spin (' J ') interactions. An extensive qualitative review of the effects of the internal motions in molecular solids is given. Part I concludes with the quantum-mechanical theory of transient experiments performed by Bloom and Norberg, Hahn and Herzog, and Proctor *et al*.

Part II provides a short but useful account of basic instrumental requirements, and is well illustrated with practical details and circuits. Quadrupole

resonance spectrometers operating up to 1000 Mc/s are described, and a previously unpublished circuit suitable for the observation of bromine echoes at 150 Mc/s is given.

The last half of the book reviews most of the experimental phenomena studied to date, and relates the observed spectra to the structures of the molecules and crystals in which the nuclei are contained. Three of these last five sections confine themselves to the much-studied halogen quadrupole spectra. The number and orientation of molecules in the unit cell, the chemical equivalence (or otherwise) of different sites for the same nucleus in the unit cell, and the changes of molecular orientation occurring during phase transitions are discussed on the basis of experimental spectra. Available quadrupole coupling data is used to discuss intermolecular binding in the solid state. The last two sections describe studies of nuclei other than halogens and the broadening of quadrupole resonance lines due to the ' change ' and ' size ' effects of impurities in the sample studied. The volume is well produced and should prove a valuable work of reference. I. J. S.

A History of Technology: Vol. IV. The Industrial Revolution, 1750–1850. (Oxford: The Clarendon Press.) [Pp. 728.] 8 guineas.

Few can have failed to see on a library shelf the great fat volumes of this series. Now we come to the great age of steam and progress. In that century chemical industry grew large, iron was smelted with coal and worked with machine tools, and the steam engine became the prime mover. New techniques in civil engineering, and agriculture, and the automation of spinning and weaving, created a new environment for the workers and new money for the entrepreneurs. Like ' the Renaissance ', there is no such thing as ' the Industrial Revolution '—only the world was never the same again. It is a fascinating question why it should have come then and there. Was it science ? Or religion ? Or the accumulation of capital ? Or a liberal political system ? Or the climate and coal of Britain ? Unfortunately, the scheme of the series does not help with answers, however tentative. Each chapter is a complete account, by an expert, of the development of one particular industry. Although the cross-references are carefully noted there are very few views of the wood. Each invention is given author and date, but one would have liked an essay or two on such general factors as the education of technicians, the working of the patent system, the problems of financing large factories, the social and economic pressure favouring innovation, and so on. Naturally, with so many authors, there is a great spread in the levels of style and interest. Sometimes there are mere catalogues, and obscure technical words used without explanation, but in general it is all quite intelligible to the layman. Reading is greatly aided by the lavish use of lucid drawings and prints. It is a beautiful piece of book production—almost a bargain at the price. J. M. Z.

Sound Pulses. By F. G. Friedlander. (Cambridge: University Press.) [Pp. xi+202.] 40*s*.

Text-books on the theory of sound deal mainly with periodic disturbances, such as harmonic wave trains and standing waves. Aperiodic disturbances receive much less attention : it is usual merely to point out that they can be resolved into harmonic components by Fourier's integral theorem. An alternative approach to aperiodic problems is based on the theory of partial differential equations of hyperbolic type and is particularly effective with disturbances having a clearly defined front. Such disturbances may be described as sound pulses : the monograph before us describes the theory of sound pulses and some of its recent developments.

The monograph is concerned throughout with the scalar wave equation of acoustics, which is a linear approximation that can only be applied to small disturbances (higher approximations are not treated here). Once the acoustic approximation is adopted, the development of the subject becomes a mathematical rather than a physical investigation. The book is therefore essentially an essay on the pulse solutions of the wave equation. Some of the work has not been published before ; for instance the appendix to chapter 3 on the focusing of acoustic shocks, the method by which the Green's function of the wedge is derived in chapter 5, and the discussion of pulse diffraction by a sphere in chapter 6.

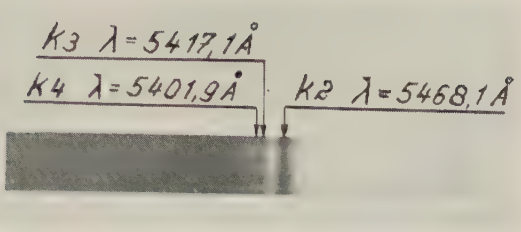
The chapter headings are : (1) Introduction ; (2) Wave fronts and characteristics ; (3) Geometrical acoustics ; (4) The application of geometrical acoustics to reflection problems ; (5) The diffraction of a pulse by a wedge ; (6) Some other diffraction problems.

After deriving the equation of acoustics in chapter 1 the author gives a brief but adequate account of the characteristics associated with the equation, and introduces influence and dependence domains. Diffraction into shadow regions, reflection and focusing are treated. These results are proved in the first place only for solutions of the wave equation with continuous derivatives. The extension to acoustic shock fronts (where the pressure is discontinuous) is made in chapter 3. It is shown that along a ray associated with an acoustic shock front the square of the pressure-jump divided by the acoustic impedance varies according to the intensity law of geometrical optics. Acoustic shocks can therefore be treated as a separate subject, *geometrical acoustics*. Geometrical acoustics can also be considered as a first approximation to a class of series solution of the wave equation, and an exposition is given of this theory. The discontinuous solutions are usually treated by first replacing the differential equation by an integral equivalent ; here the author gives an alternative treatment by L. Schwartz's theory of distributions, a brief exposition of which is given. In the next chapter the theory is extended to reflected pulses. Among other examples the reflection of a plane pulse by a convex paraboloid is solved in closed form and also by a series. Chapter 5 is devoted to explicit solutions of wedge problems. The theory of distributions is used again in an appendix. Chapter 6 treats diffraction by a circular cylinder and a sphere and other problems. Here Bessel functions of complex order appear, as might have been expected from the analogous time-periodic problem. Two sections are devoted to stratified media, bounded by a plane.

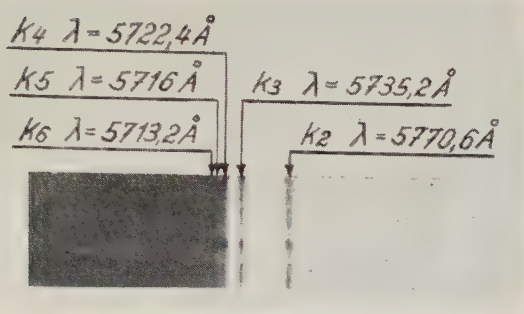
In a short monograph of 200 pages there can be little room for lengthy explanations, but the arguments though concise are complete. The mathematical treatment is elegant and appropriate. The style is remarkable for the author always says exactly what he means. Altogether the author deserves warm congratulations on this fine piece of work. F. U.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

Fig. 2



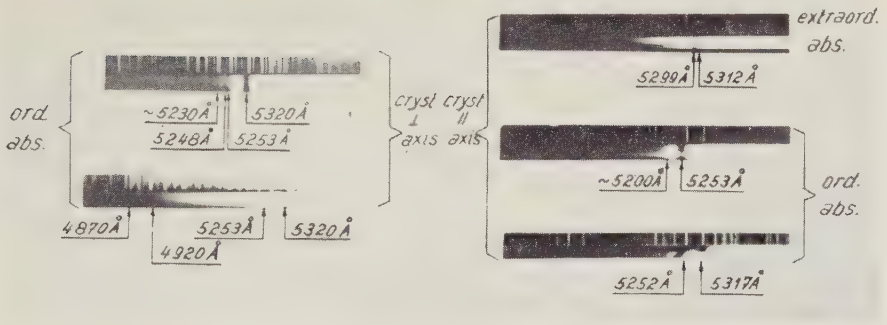
(*a*)



(*b*)

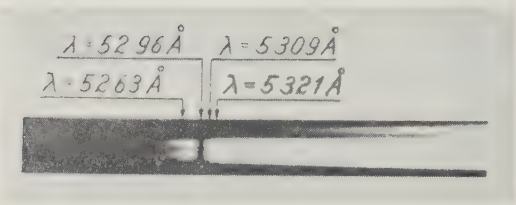
Spectra of the yellow and green series of Cu_2O at 4·2°K.

Fig. 5



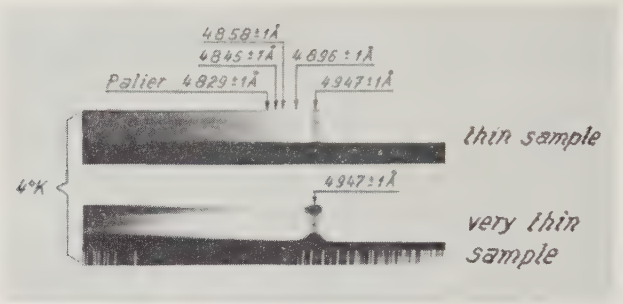
Ordinary and extraordinary spectra of HgI_2 at 4·2°K in thin crystals.

Fig. 6



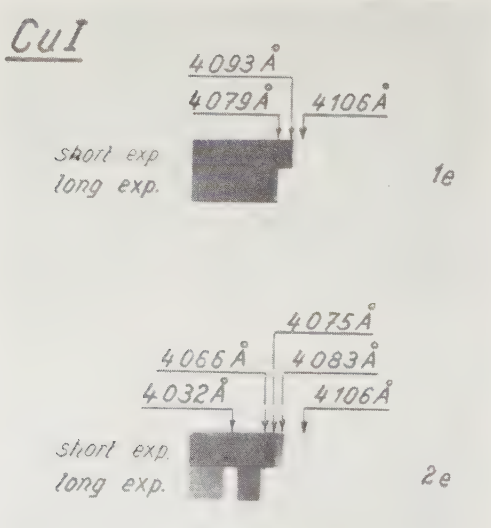
Extraordinary spectrum of HgI_2 at 4·2° K showing a group of three lines observed with a thick crystal.

Fig. 8



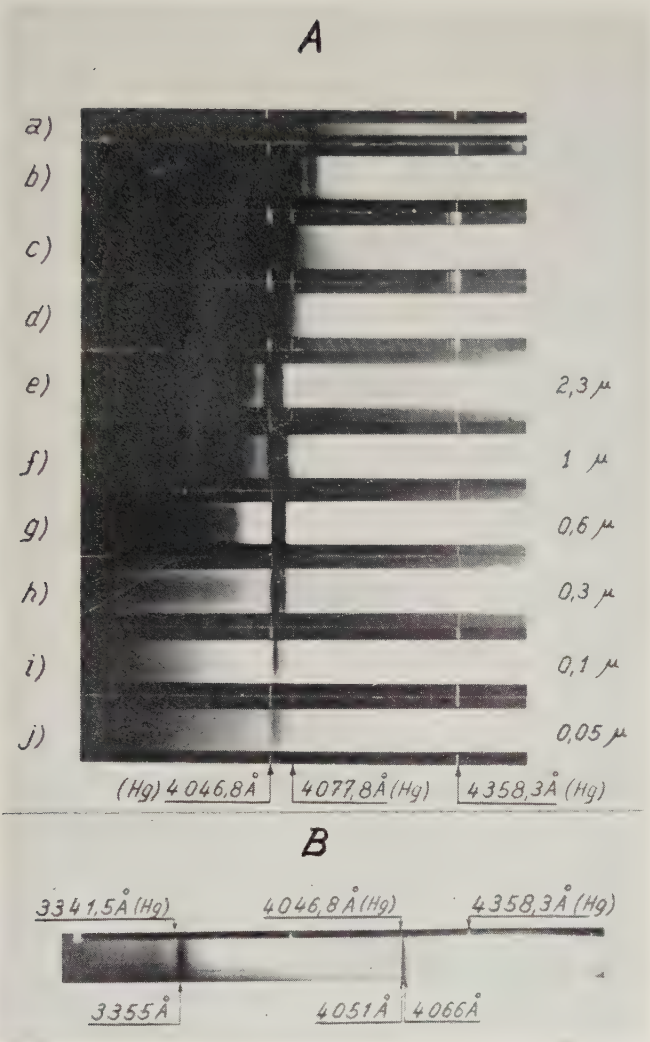
Spectra of PbI_2 obtained at 4·2°K with a moderately thin and a very thin microcrystalline film.

Fig. 10



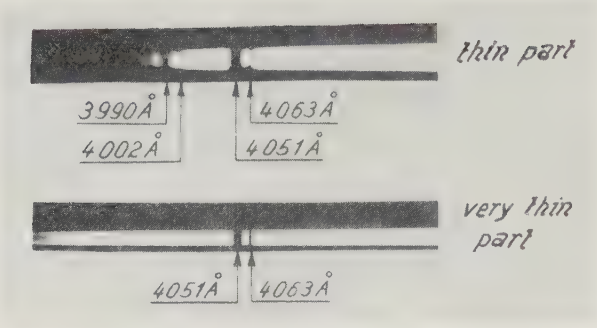
Spectra of CuI at 4·2°K. Specimens 1 e and 2 e (structure sensitive or irregular lines).

Fig. 12



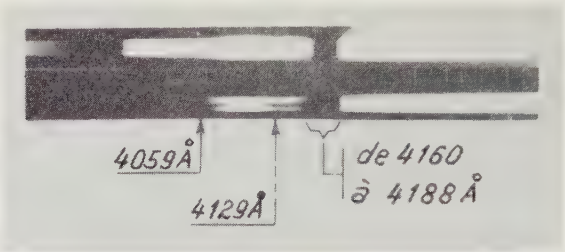
(*A*) Spectra of a CuI specimen at 77·3°K observed at different thicknesses in the specimen. The 'raies ultimes' are observed in the thin part of the specimen. (*B*) 'Raies ultimes' of CuI both in the visible and in the ultra-violet.

Fig. 13



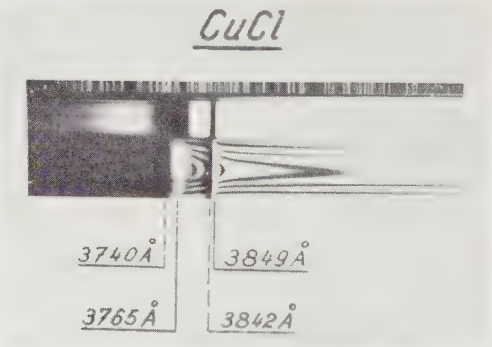
First and second doublet (the last reinforced in the reproduction) of the series of CuI and the ' raies ultimes ' observed with a very thin sample, both at 4·2°K.

Fig. 14



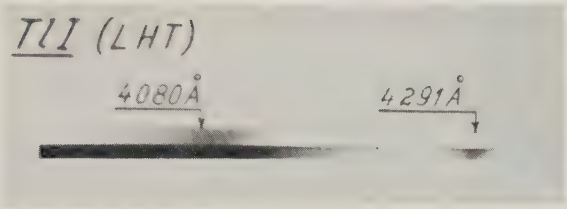
Spectra of a thin and of a thicker part of a specimen of CuBr at 4·2° K. The spectrum has no relation to the absorption curve observed by Fesefeldt.

Fig. 15



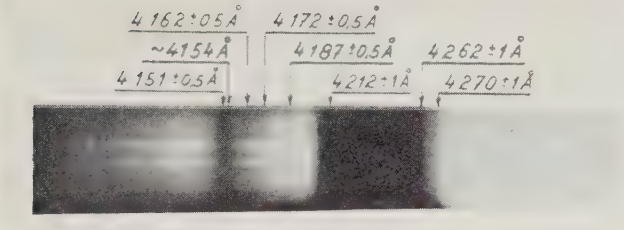
' Raies ultimes ' of CuCl at 4·2°K (upper spectrum) and the reflection spectrum of the same specimen showing residual and ' missing ' rays and interference fringes. The sample was in the form of a meniscus.

Fig. 16



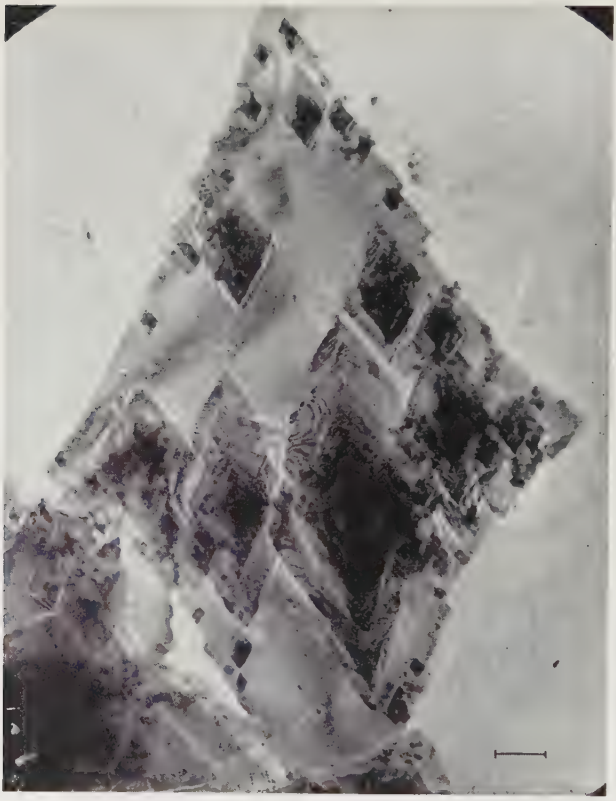
Spectra of TlI at 4·2°K (thin and thicker specimens).

Fig. 17



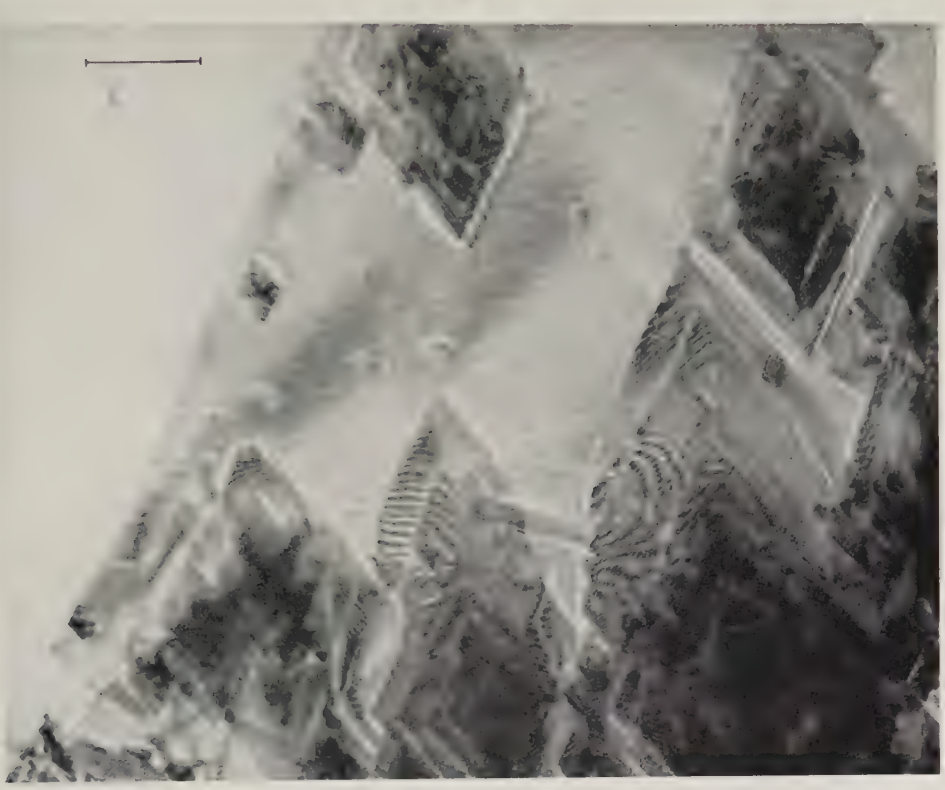
The absorption spectrum of γ-AgI at 4·2°K.

Fig. 1



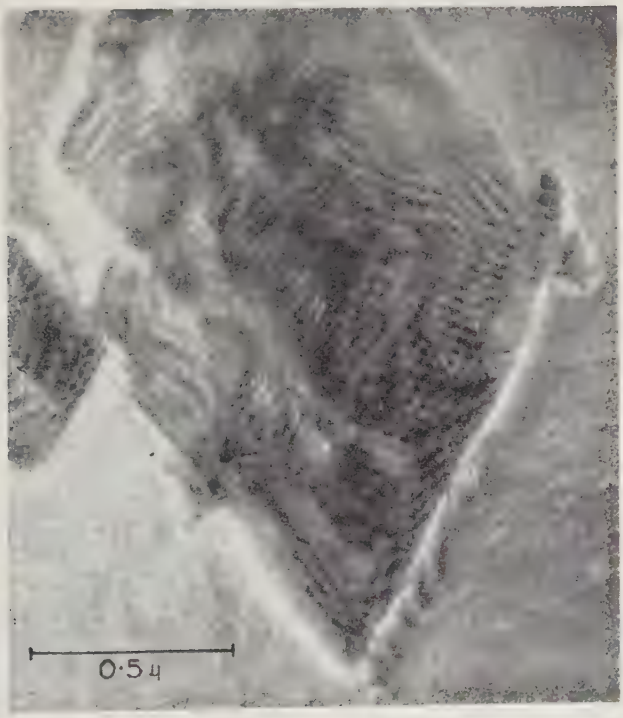
Polyethylene crystal with moiré patterns. ×6500.

Fig. 2



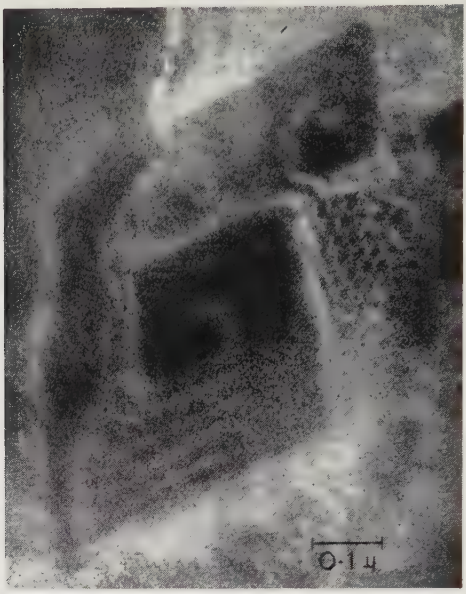
Detail of fig. 1. ×15 000.

Fig. 3



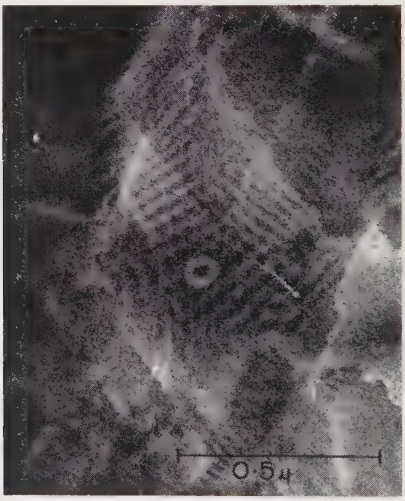
Detail of fig. 2, showing the resolved pseudo-image of the lattice and fringes perpendicular to the nearest {110} faces. ×53 200.

Fig. 4



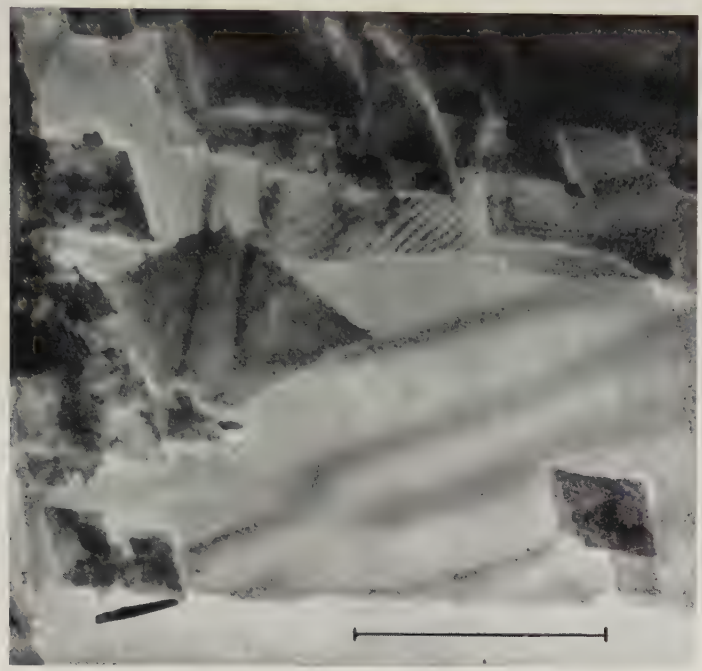
Detail of a polyethylene crystal showing the resolved pseudo-image of the lattice and distortions in the fringes around a screw dislocation. ×91 000.

Fig. 5



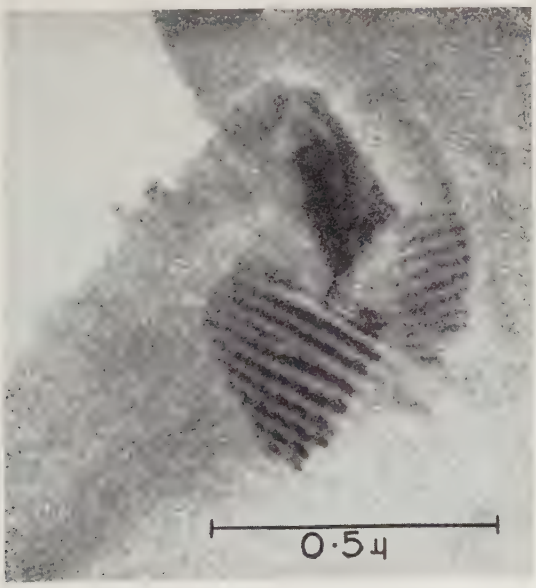
Detail of a polyethylene crystal showing the resolved pseudo-image of the lattice and fringes perpendicular to {110} faces belonging to adjacent quadrants. ×53 200.

Fig. 6



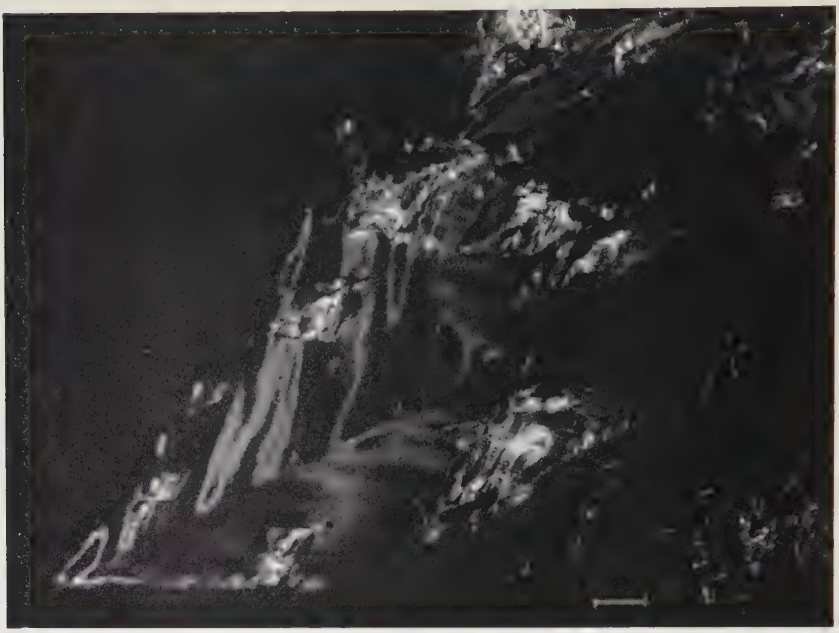
Detail of a polyethylene crystal with broad moiré fringes along [010], situated between Bragg extinction fringes. ×33 600.

Fig. 7



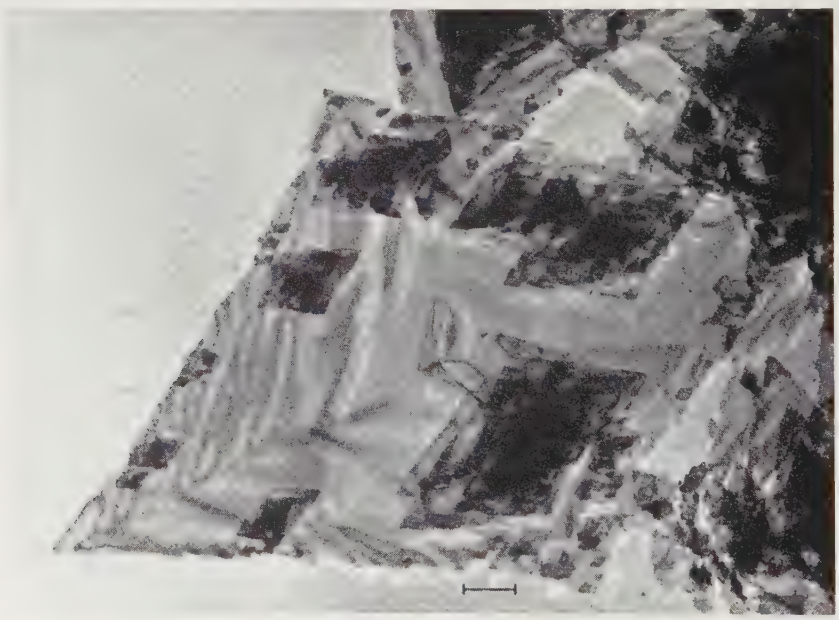
Detail of a polyethylene crystal showing a dislocation. ×76 000.

Fig. 8



Polyethylene crystal in dark ground illumination. ×6650.

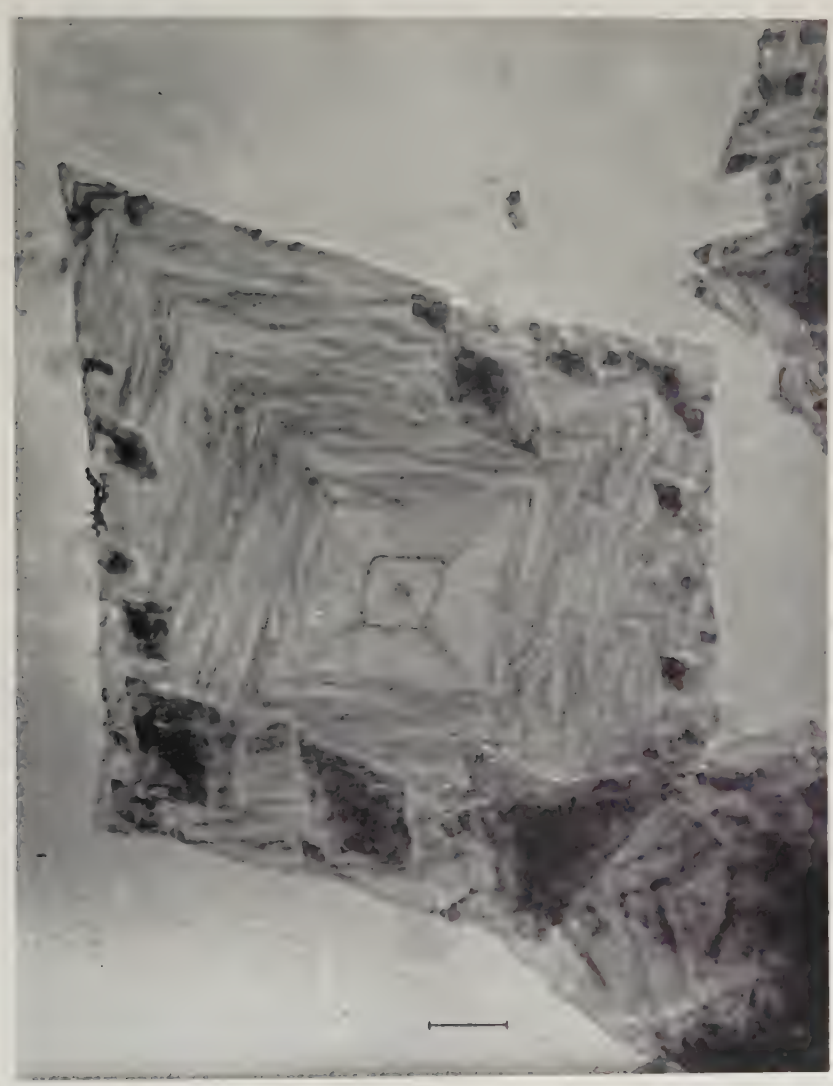
Fig. 9



The same crystal as in fig. 8 in bright field. ×6650.

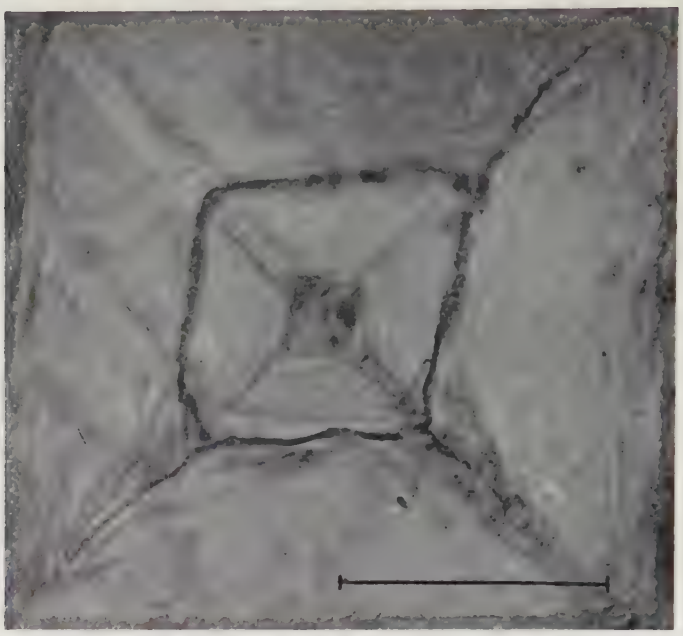
Note the white ghost images parallel to the Bragg extinction fringes and to parts of the crystal faces. Note in particular the opposite direction of the displacements of the two D shaped ghost images with respect to the corresponding extinction fringes in the centre. Also note that the Bragg extinction fringes in fig. 8 are visible only in one half of the left side of the crystal, while they are present in both halves in fig. 9.

Fig. 10



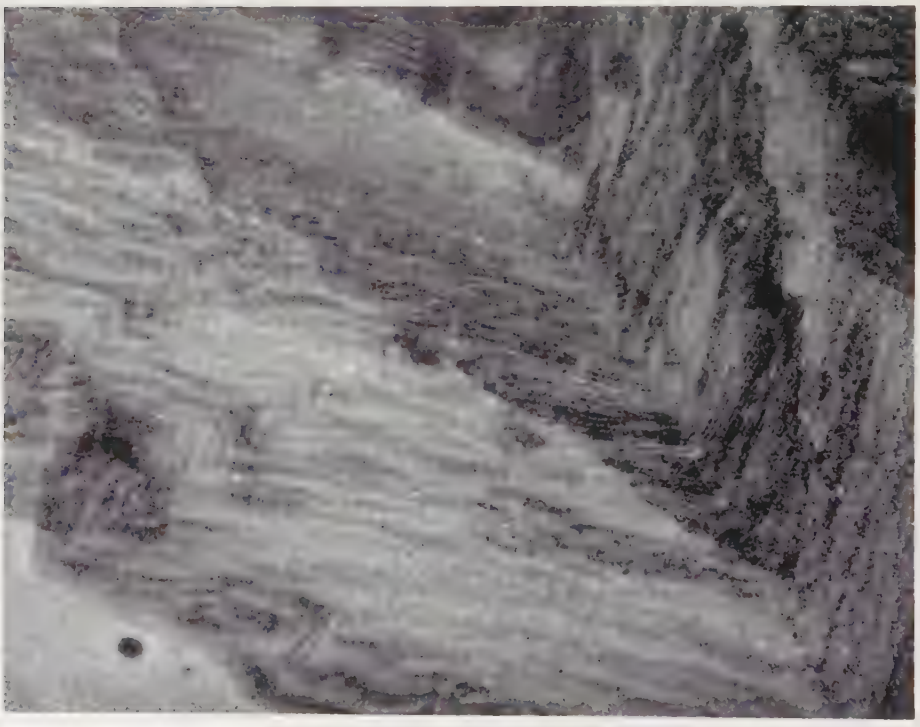
Polyethylene crystal photographed immediately after drying, showing a structure within single layers, dividing the crystal in four quadrants. $\times 10\,000$.

Fig. 11



Central detail of fig. 10, showing corrugation. ×35 000.

Fig. 12

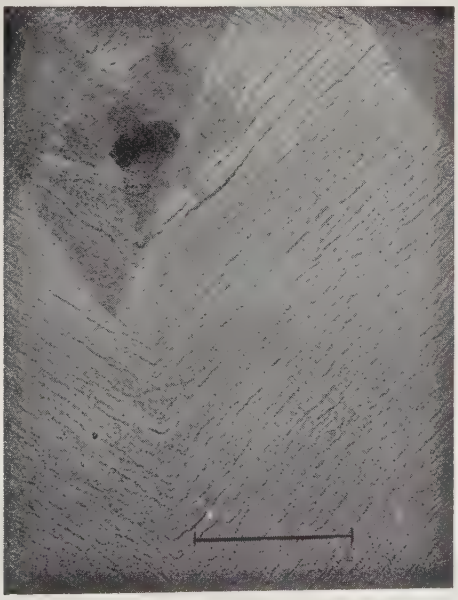


Detail of a polyethylene crystal of the type as in fig. 10, showing fault bands due to twinning and fine moiré fringes perpendicular to {110} faces in adjacent quadrants. ×37 000.

Fig. 13

Detail of a polyethylene crystal with sharp Bragg extinction fringes in single layers. ×18 000.

Fig. 14

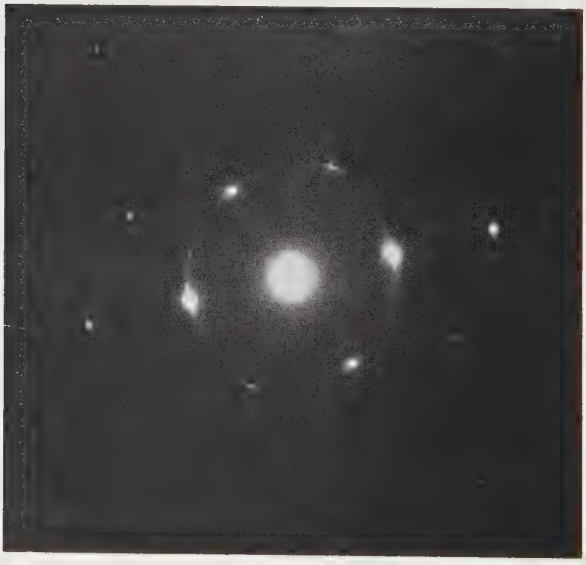


Detail of a polyethylene crystal showing surface corrugations in the absence of extinction effects, changing direction along the crystal diagonals. ×21 000.

Fig. 15

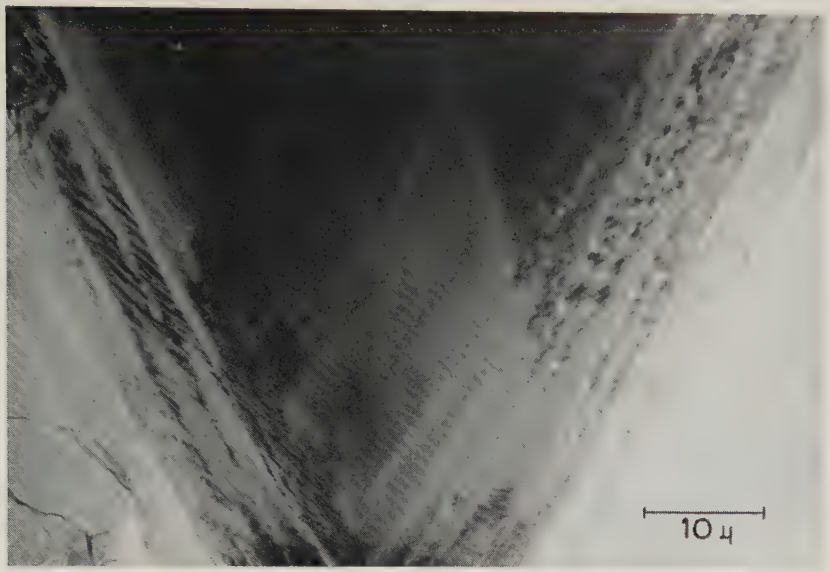
Polyethylene crystal from a lower molecular weight fraction showing a fourfold leaf pattern. Shadowed. ×4 900.

Fig. 16



Electron diffraction pattern from the edge of a crystal as in fig. 15.

Fig. 17



Part of a crystal as in fig. 15, with moiré fringes perpendicular to {110} faces of adjacent quadrants.

Fig. 18

Electron diffraction pattern of superimposed twins with secondary reflections near the centre of the pattern.

Fig. 19

(*a*)

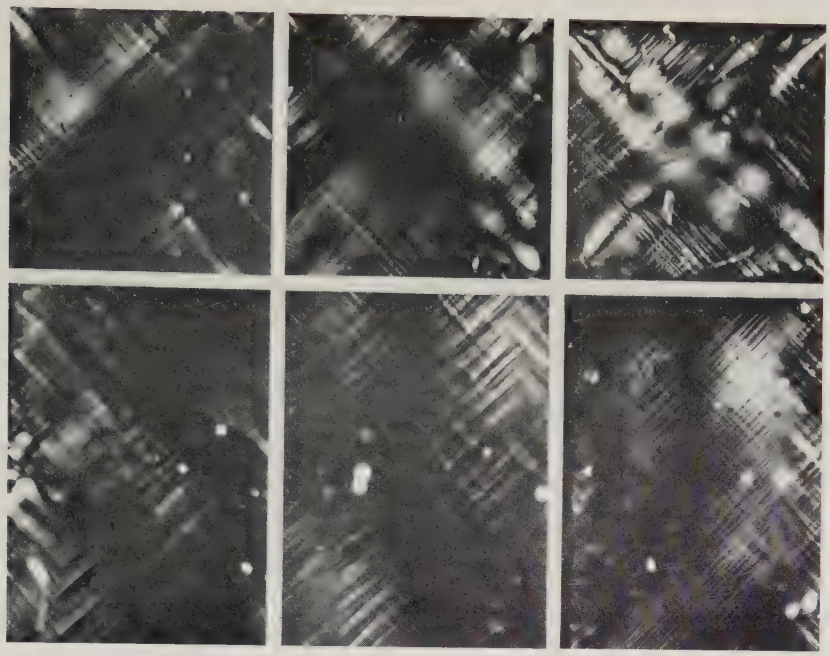
(*b*)

Electron diffraction patterns with secondary diffraction effects due to rotated, superimposed layers.

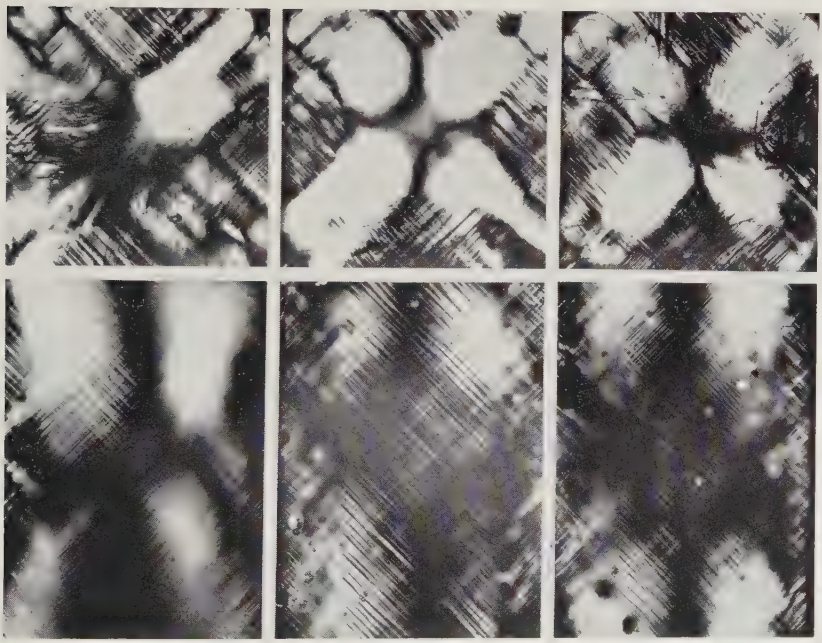
(*a*) Lightly printed, showing asymmetric spikes around the main reflections.

(*b*) Heavily printed, showing spikes around the centre of the pattern.

Fig. 1



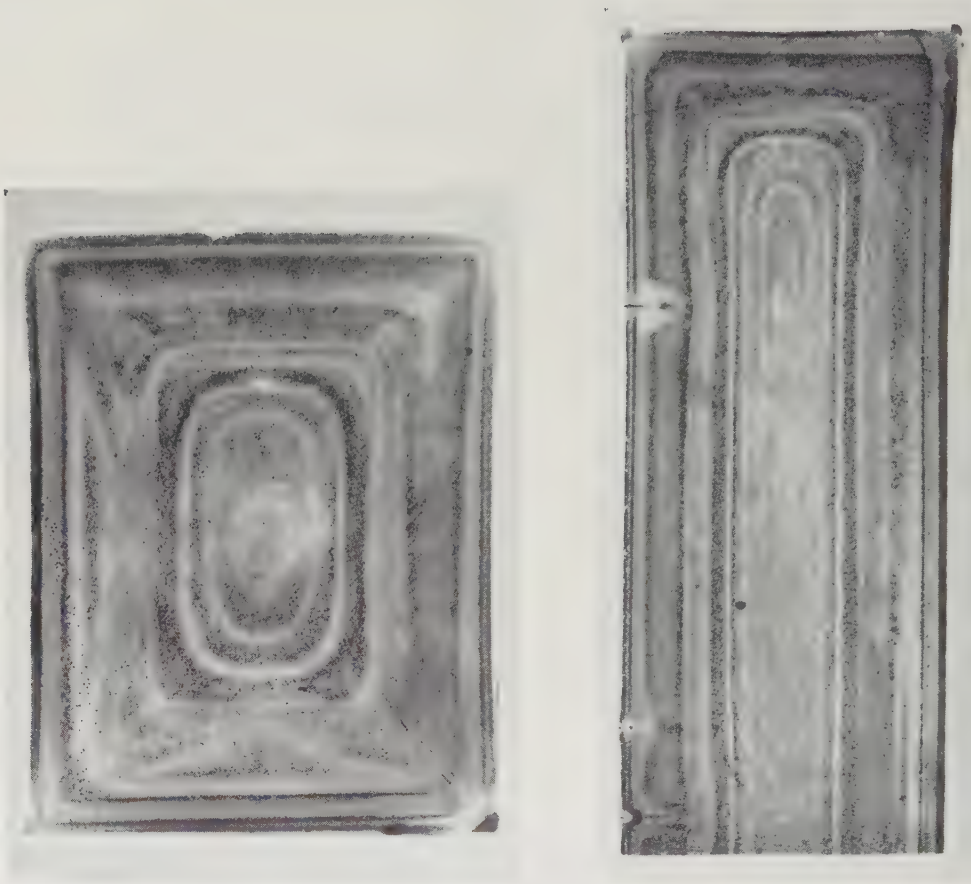
(*a*) Still Air (*b*) Air Quench (*c*) Liquid Air



(*d*) CCl_4 (*e*) Silicone Oil (*f*) Brine

Birefringent patterns seen in end and side faces of quenched crystals. Order (*a*) to (*f*) is increasing severity of quench. (×12·5.)

Fig. 9



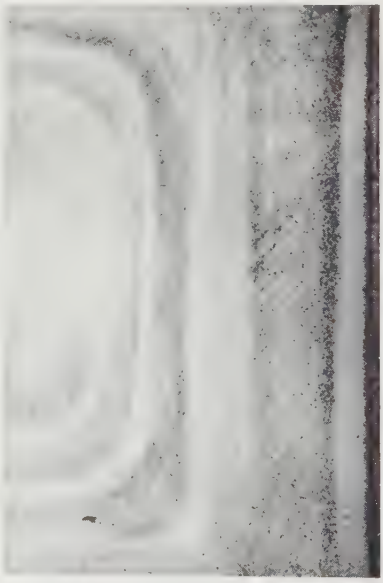
(*a*) Transverse section. (×35.) (*b*) Longitudinal section. (×28.)

Longitudinal and transverse sections of a decorated quenched crystal.

Fig. 10

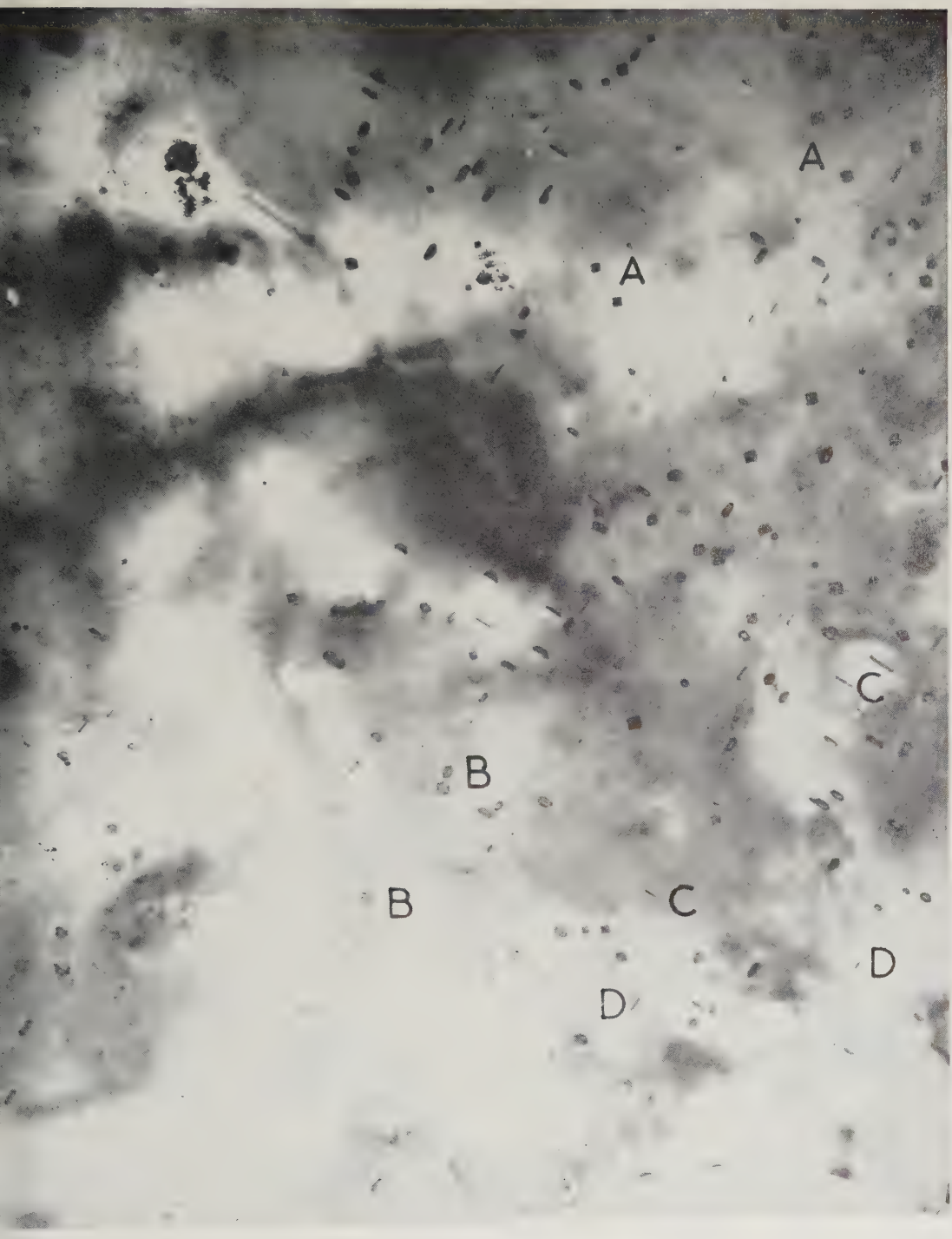
Enlarged view of longitudinal section of fig. 9 (*b*). (×72.)

Fig. 11



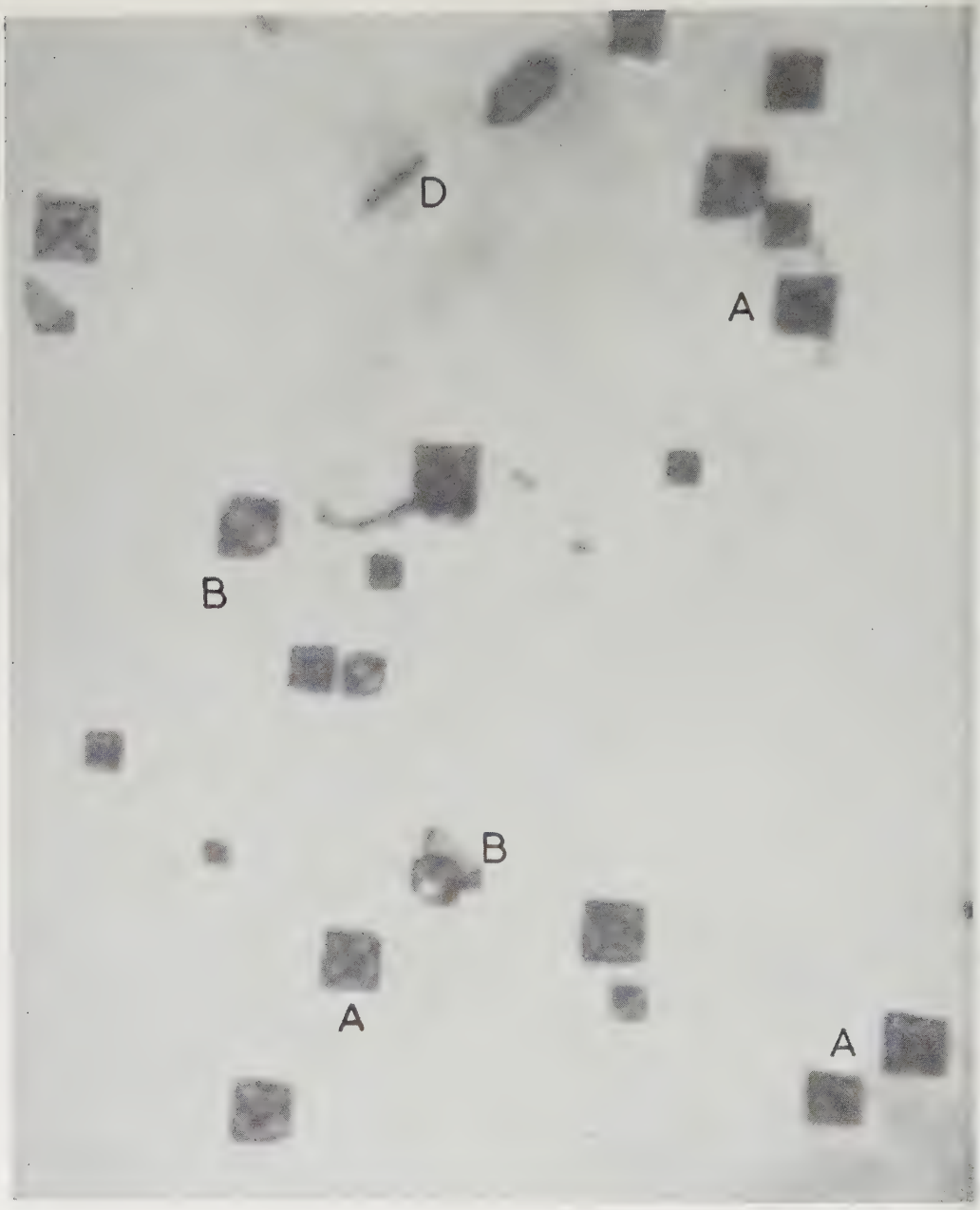
Enlarged view of transverse section of fig. 9 (*a*). (×72.)

Fig. 1



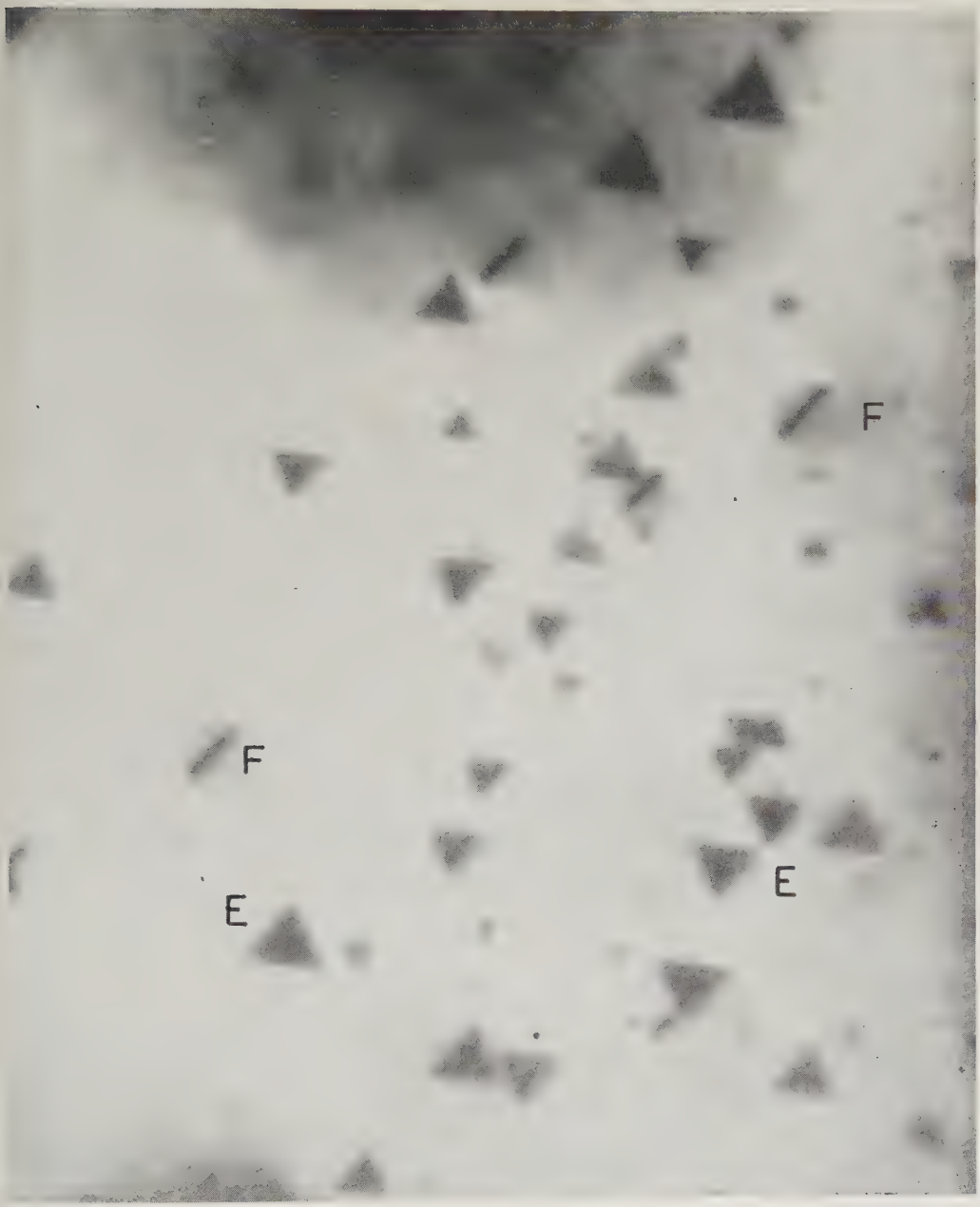
fects in polycrystalline gold quenched from about 960°c into iced brine and aged for one hour at 100°c. Notice square (A), elongated hexagon (B), and line (C, D) shapes. The normal to the foil is parallel to [100]. Mag. ×57 000.

Fig. 2



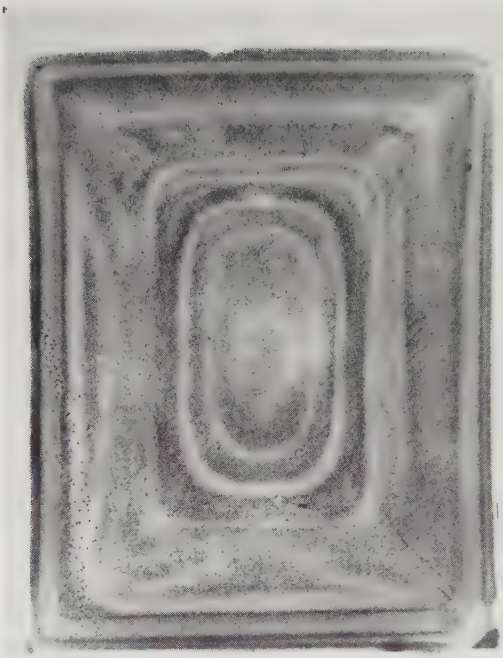
High magnification micrograph of a specimen with the normal to the foil in the [100] orientation. Note the square (A), elongated hexagon (B), and line (C, D) shapes. Mag. ×186 000.

Fig. 3

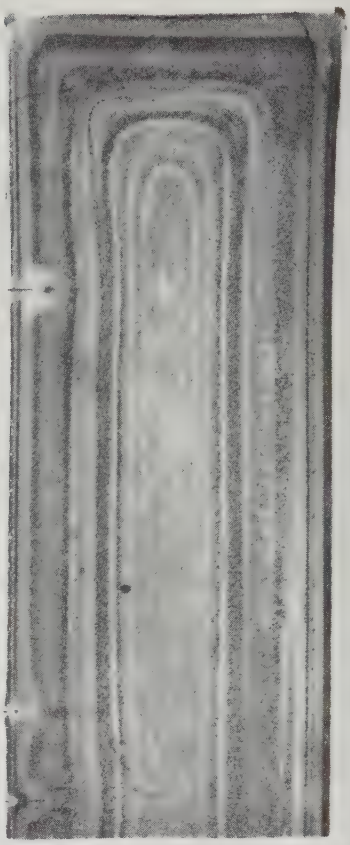


High magnification micrograph of a specimen with the normal to the foil in the [110] orientation. Note the triangle (E) and line (F) shapes. Mag. × 255 000.

Fig. 9



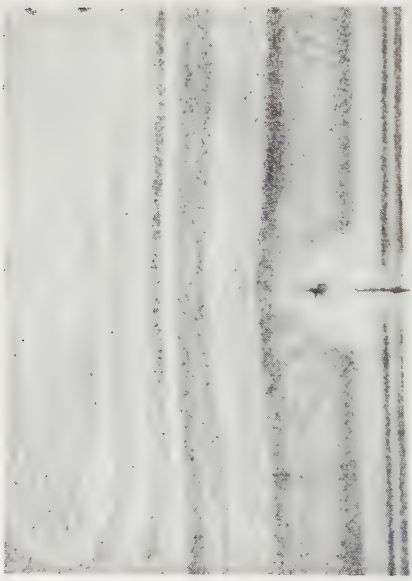
(*a*) Transverse section. (×35.)



(*b*) Longitudinal section. (×28.)

Longitudinal and transverse sections of a decorated quenched crystal.

Fig. 10

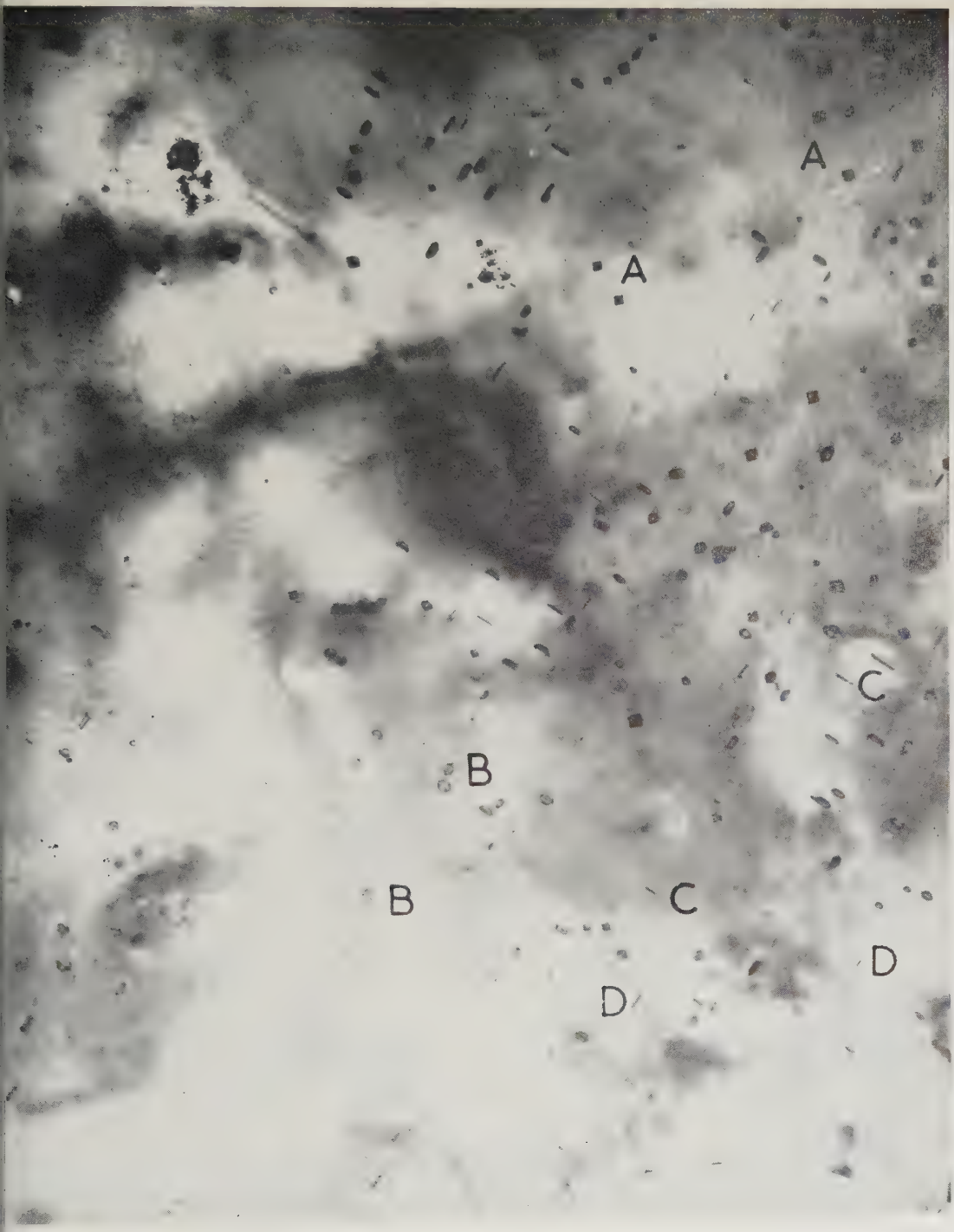


Enlarged view of longitudinal section of fig. 9 (*b*). (×72.)

Fig. 11

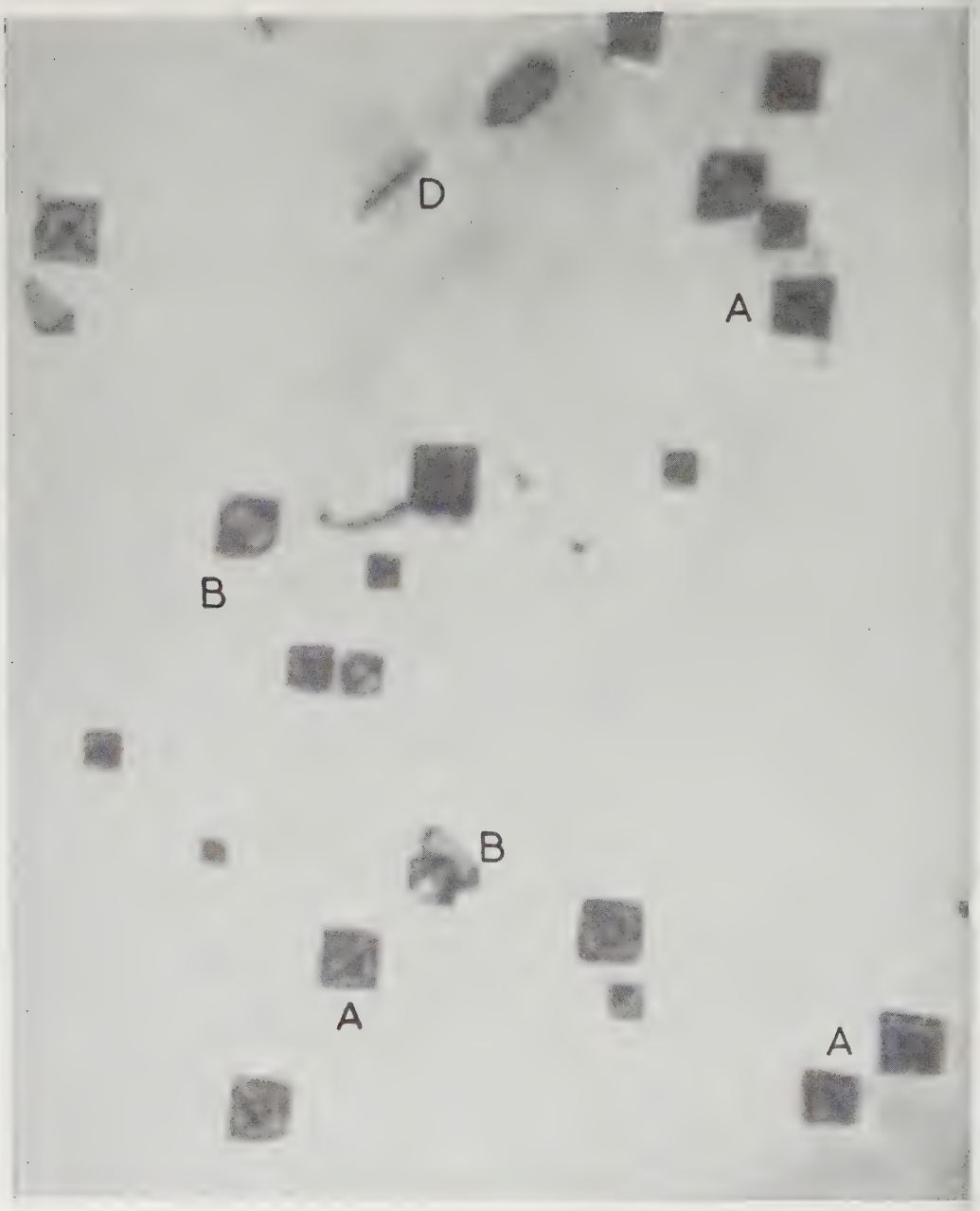
Enlarged view of transverse section of fig. 9 (*a*). (×72.)

Fig. 1



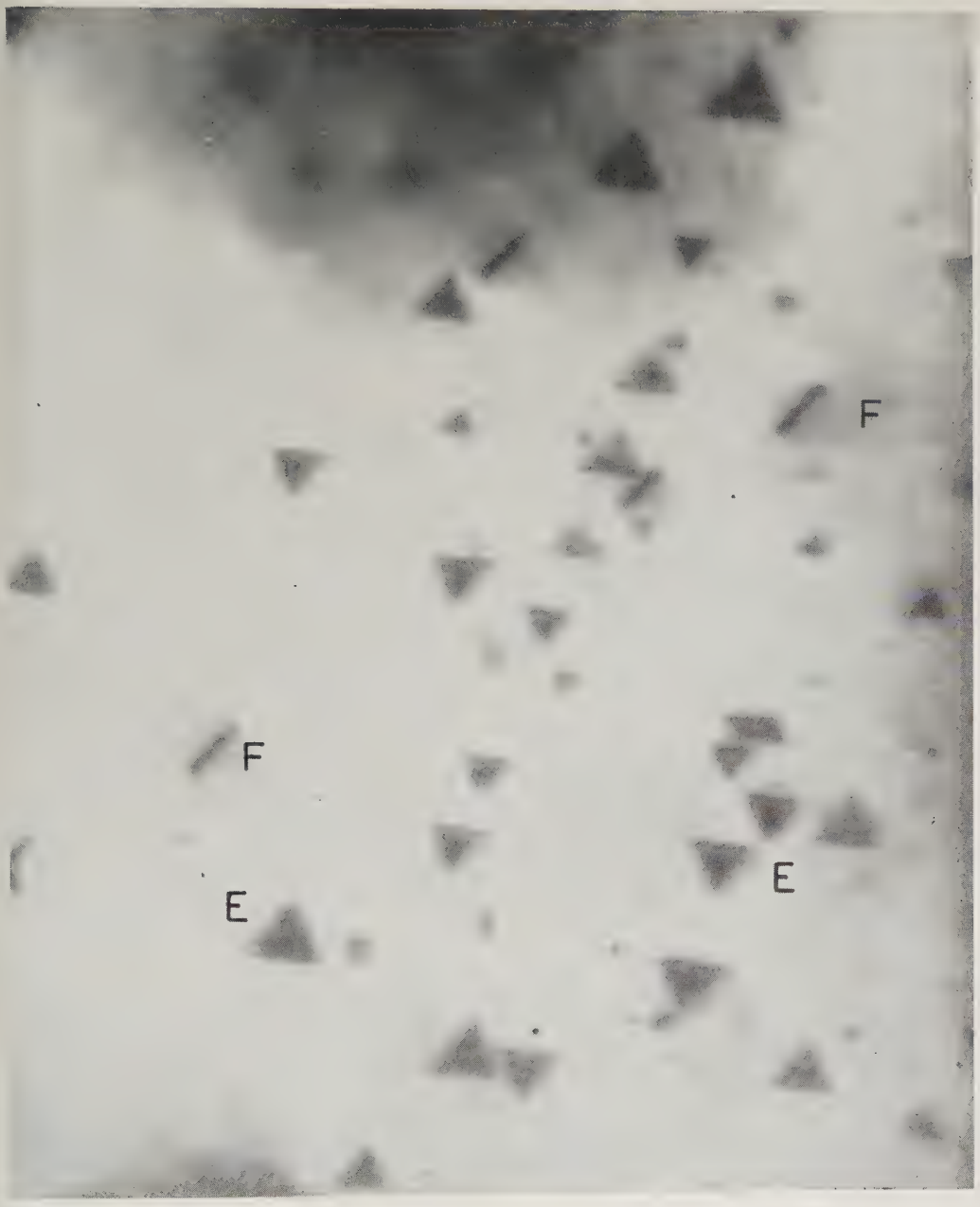
fects in polycrystalline gold quenched from about 960°c into iced brine and aged for one hour at 100°c. Notice square (A), elongated hexagon (B), and line (C, D) shapes. The normal to the foil is parallel to [100]. Mag. ×57 000.

Fig. 2



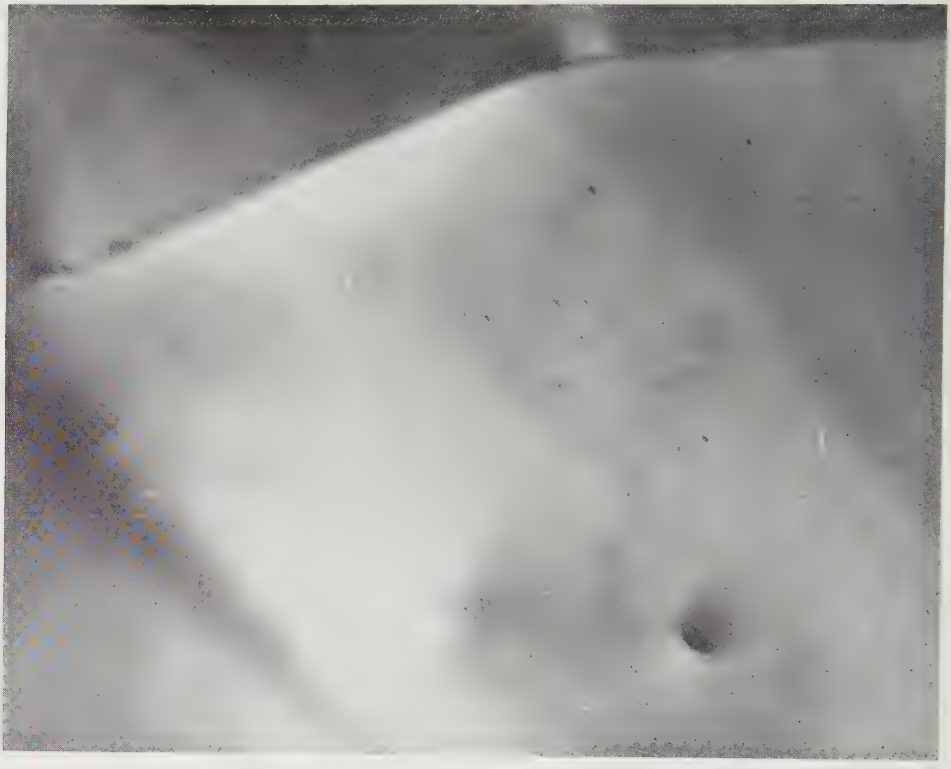
High magnification micrograph of a specimen with the normal to the foil in the [100] orientation. Note the square (A), elongated hexagon (B), and line (C, D) shapes. Mag. ×186 000.

Fig. 3



High magnification micrograph of a specimen with the normal to the foil in the [110] orientation. Note the triangle (E) and line (F) shapes. Mag. × 255 000.

Fig. 4



Electron transmission micrograph of a slowly cooled specimen of gold. The line across the micrograph is a grain boundary. Mag. × 40 000.

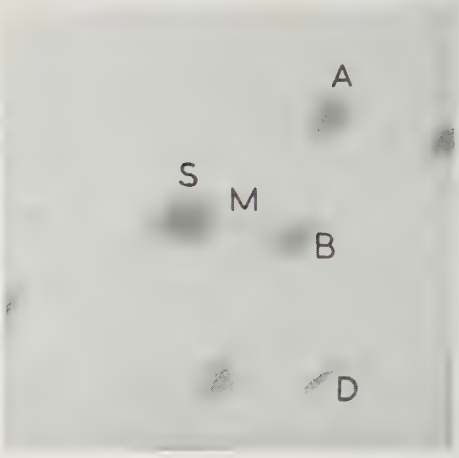
A micrograph of a specimen in the [110] orientation showing fringes parallel to $[1\bar{1}0]$ or AC, this being the [110] direction lying in the plane of the foil. The contrast is redrawn in fig. 6 (*b*). Mag. $\times$ 180 000.

Fig. 7

1000 Å

The type of contrast observed when the tetrahedron is viewed along a [100] direction. Notice the regular sets of dots. The directions of the edges of the square correspond to the projections of the [110] directions not in the plane of the foil and the diagonals of the square correspond to the [110] directions actually in the plane of the foil. Mag. × 450 000.

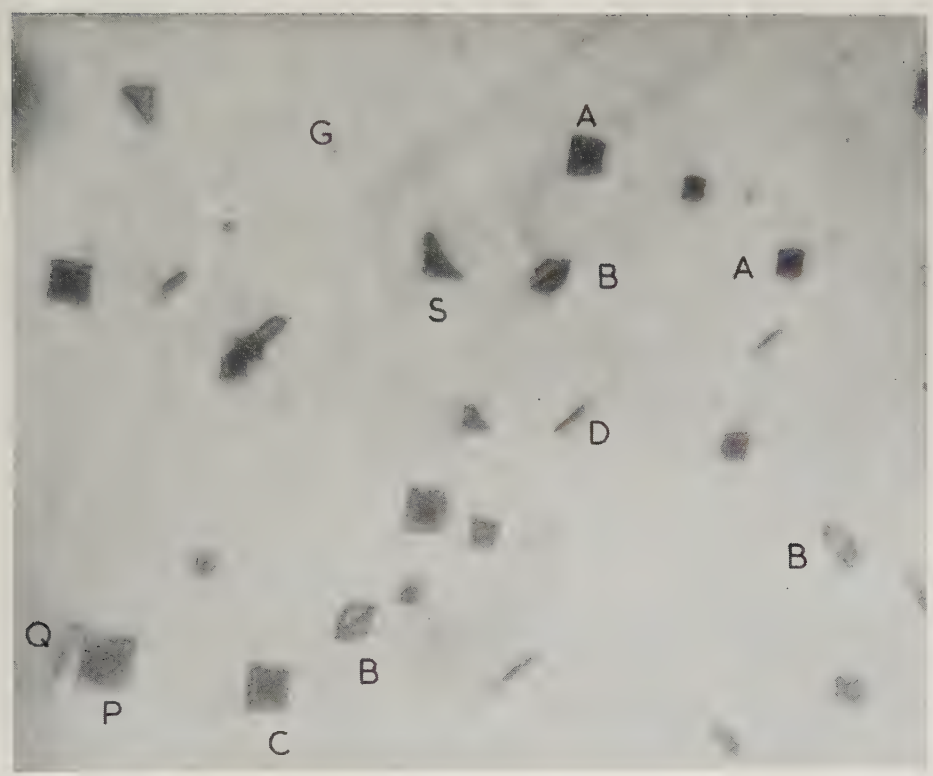
Fig. 9 (*a*)



1000 Å

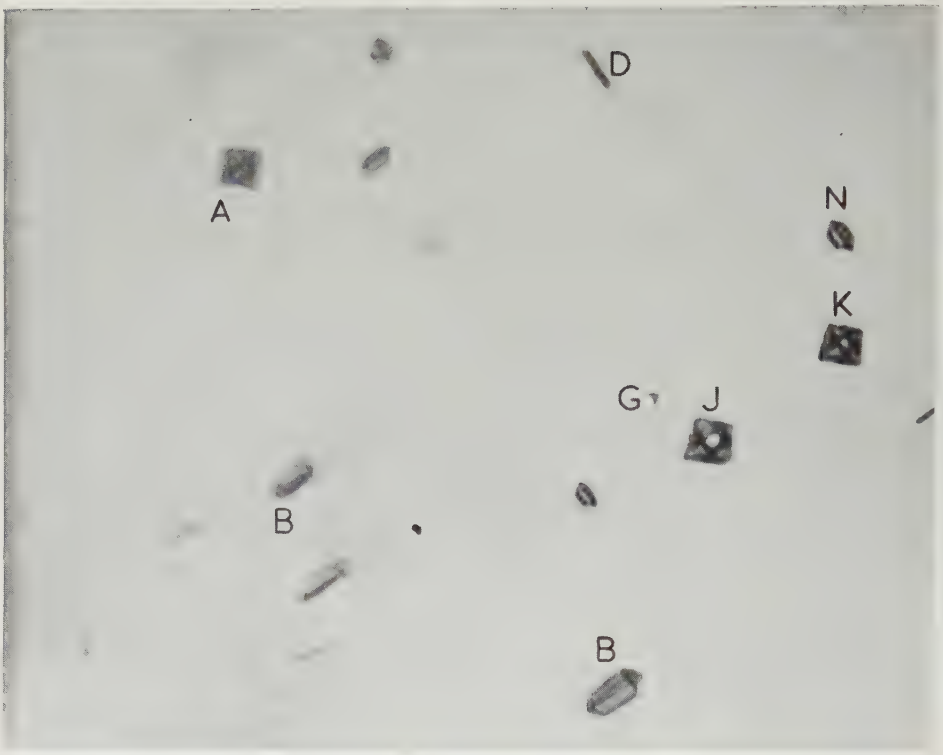
A dislocation M is seen near a complete defect S. Mag. × 170 000.

Fig. 9 (*b*)



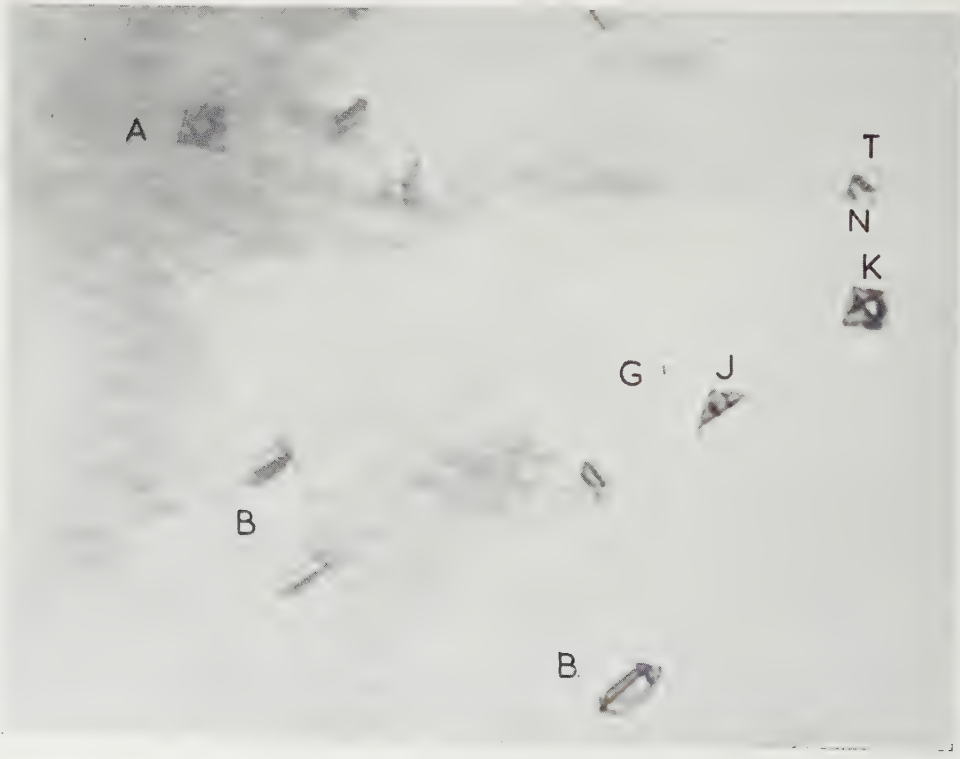
A micrograph taken after fig. 9 (*a*), showing the collapse of the defect S and the cross-slip trace of the dislocation M. Notice the distortion of the defect P by the dislocation Q. Mag. × 170 000.

Fig. 10 (*a*)



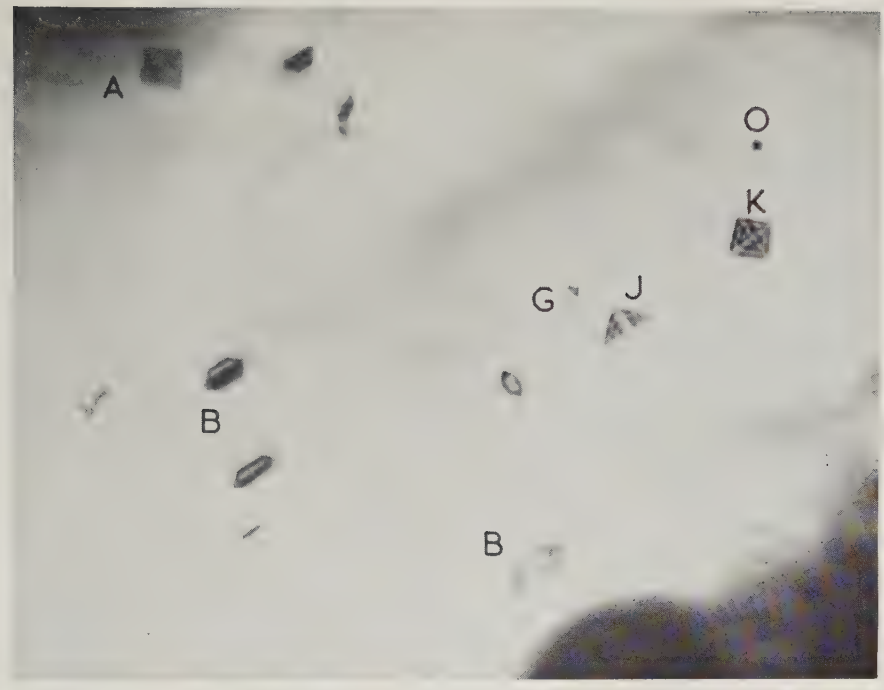
1 000Å

Fig. 10 (*b*)



1000 Å

Fig. 10 (*c*)



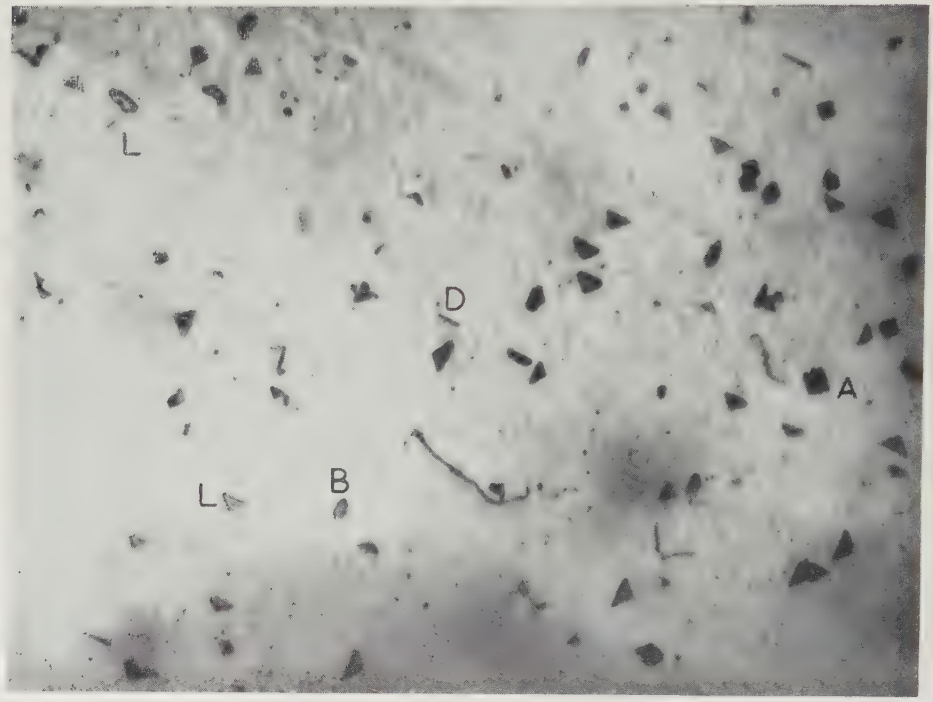
1000 Å

(*a*) The first of a sequence of micrographs taken with a foil in the [100] orientation. Notice the squares J, K, and the truncated tetrahedron N. Mag. × 170 000.

(*b*) In this micrograph the defect J has collapsed. Notice also the slip trace left by a dislocation T interacting with the truncated tetrahedron N. Mag. × 170 000.

(*c*) This micrograph shows the reformation of the defect J, and also shows that the dislocation T has disappeared leaving a small but complete tetrahedron O. Mag. × 170 000.

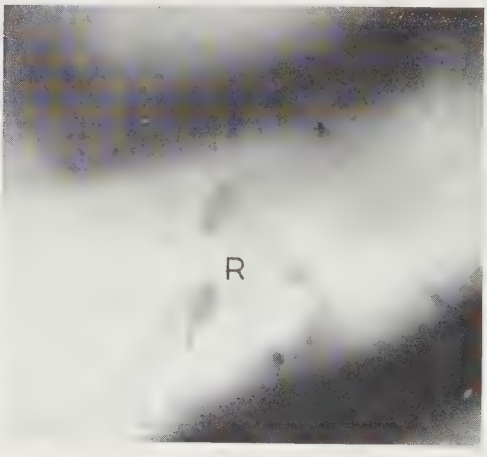
Fig. 12



1000 Å

A micrograph of an area in the [100] orientation showing triangles and also apparently glissile loops, L. Mag. × 76 000.

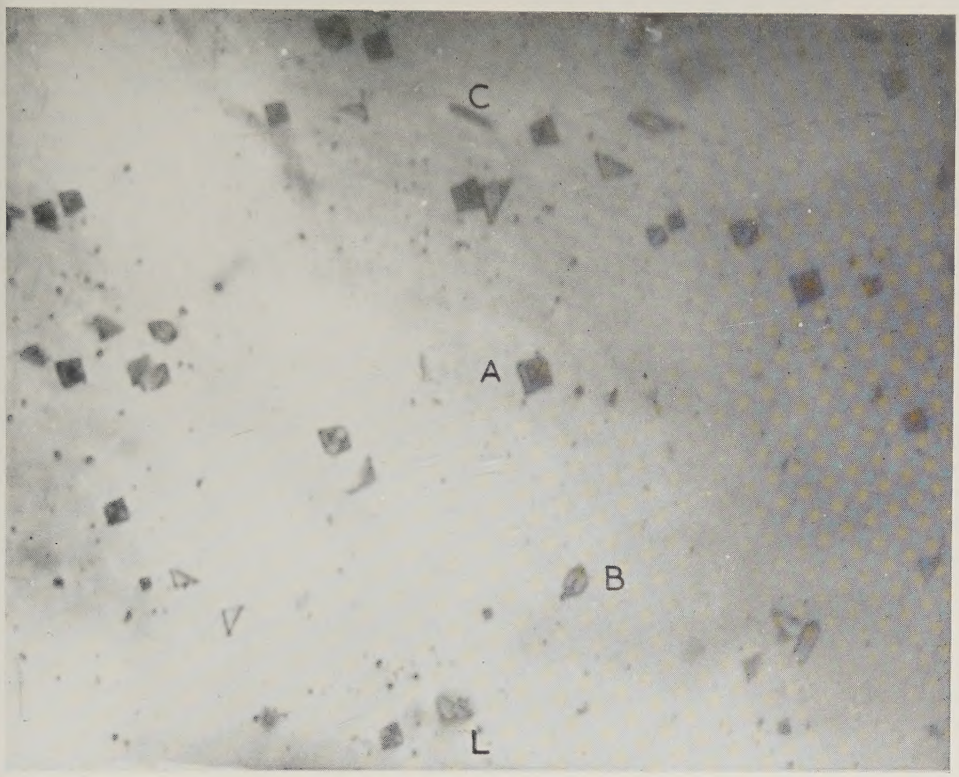
Fig. 14



1000 Å

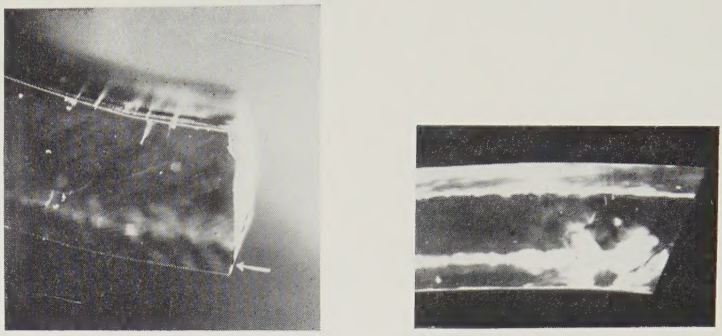
A micrograph showing the emergence of a glissile dislocation from a defect R. Mag. × 160 000.

Fig. 13



A micrograph of an area in the [100] orientation showing effects similar to those on fig. 12. Mag. × 180 000.

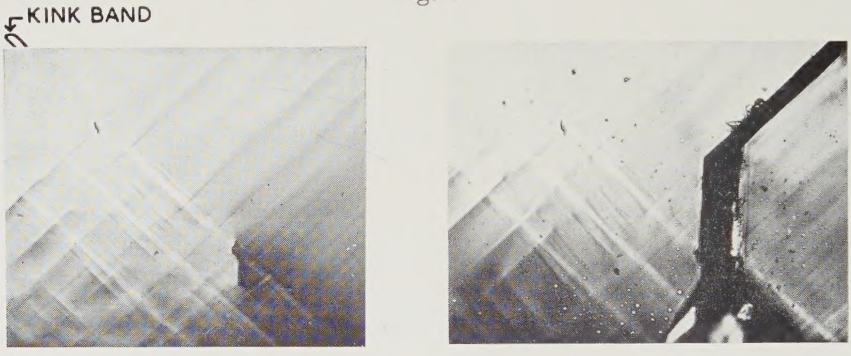
Fig. 1



(*a*) (*b*)

Slits formed by bending magnesium oxide single crystals. (*a*) Slits in the compression region of a fractured crystal (×10). (*b*) Slits in the tension region of an unfractured crystal (×6).

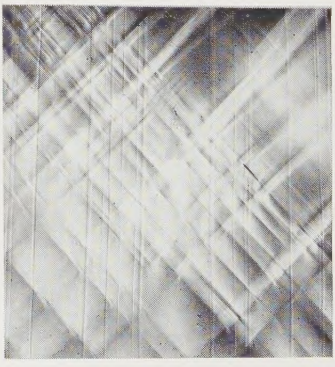
Fig. 2



(*a*) (*b*)

The orientation and situation of the slits shown in fig. 1 (*b*). (*a*) Stable Stroh-cracks lying along the edge of a kink-band (polarized illumination ×150). (*b*) Path of final fracture through the crystal (polarized illumination ×150). Tension direction horizontal in both photomicrographs.

Fig. 3



Typical Stroh-crack formed during bending (polarized illumination ×500).

Fig. 2